

Univerzita Karlova v Praze

Přírodovědecká fakulta



**Porovnání použitelnosti bibliografických databází
pro řešení biochemických problémů**

**Comparison of efficiency of bibliographic databases
for biochemical problems**

Bakalářská práce

Praha 2010

Řešitel: Dagmar Frantíková

Školitel: prof. RNDr. Jiří Hudeček, CSc.

Prohlašuji, že jsem tuto bakalářskou práci vypracovala samostatně pod vedením školitele prof. RNDr. Jiřího Hudečka, CSc. a všechny použité prameny jsem řádně citovala.

Místo a datum:

Podpis:

Poděkování

Tímto bych chtěla poděkovat prof. RNDr. Jiřímu Hudečkovi, CSc. za odborné vedení, trpělivost a ochotu při konzultacích a čas, který mi věnoval. Dále bych ráda poděkovala své rodině za podporu při psaní práce.

Abstrakt:

Práce se zabývá porovnáním čtyř v biochemii hojně využívaných databází: placených SCOPUS a Web of Science a neplacených MEDLINE a Google Scholar. Byla hodnocena jejich účinnost a vhodnost pro řešení biochemických problémů. Byly zvoleny tři modelové dotazy, které byly zadávány do databází: „ellipticine“, „Lowry method“ a „serine racemase“. Vyhledané záznamy byly poté řazeny dle předem určených kritérií podle relevance.

Jako nejlepší vychází ze srovnání placená databáze SCOPUS, a to díky přehlednému rozhraní a vysokému podílu relevantních vyhledaných článků ve srovnání s databázemi ostatními. Velmi dobře se také osvědčila databáze MEDLINE, jenž je neplacená a při nedostupnosti databáze předplacené by při hledání pravděpodobně stačila. Nejlepší výsledky byly nalezeny po kombinaci těchto dvou. I Web of Science poskytla dobré výsledky, avšak ve srovnání s databázemi ostatními se nezdála jako nejužitečnější. Jako nejhorší jsem shledala Google Scholar, který bych doporučila spíše pro hledání informací o základních, ne příliš specializovaných problémech. Databáze totiž vyhledá velké množství relevantních článků, avšak je nutné záznamy hledat mezi nemalým počtem nerelevantních. Práce je tedy poté daleko časově náročnější.

Klíčová slova:

bibliografické databáze, SCOPUS, Web of Science, Web of Knowledge, PubMed, MEDLINE, Google Scholar, ellipticin, Lowryho metoda, serinová racemasa

Abstract:

The study presents a comparison between four bibliographic databases widely used in biochemistry: paid SCOPUS and Web of Science and free-accessible MEDLINE and Google Scholar. Their efficacy and suitability for solving biochemical problems was evaluated. Three model problems were chosen and entered to databases: „ellipticine“, „Lowry method“ and „serine racemase“. The located articles in each database were then sorted by their relevance.

The best results were obtained with the SCOPUS database (highest proportion of relevant results; thus database has at the same time user-friendly interface). Very good was also free MEDLINE database and if paid databases are unavailable, MEDLINE would be the right option. The best results were found with a combination of both above named databases. Also database Web of Science gave good and relevant results but was not as useful as SCOPUS or MEDLINE. The worst results in this study was given by the database Google Scholar. It would be recommended for searches of basic, not so specific problems. This database locates many articles but with questionable proportion of relevant articles – and these have to be found manually, which would be time-consuming. Thesis in Czech.

Keywords:

bibliographic databases, SCOPUS, Web of Science, Web of Knowledge, PubMed, MEDLINE, Google Scholar, ellipticine, Lowry method, serine racemase

Obsah:

Úvod a cíl práce.....	7
1. Databáze jako zdroj vědeckých informací.....	8
1.1. Srovnávání databází.....	8
1.1.1. Faktografické databáze a jejich porovnání.....	8
1.1.2. Bibliografické databáze a jejich porovnání.....	11
1.2. Charakteristika jednotlivých databází.....	13
1.2.1. SCOPUS.....	13
1.2.2. Web of Knowledge.....	15
1.2.3. Rozhraní PubMed databáze MEDLINE.....	18
1.2.4. Google Scholar.....	19
2. Metody vyhledávání.....	22
2.1. Vyhledávání pojmu „ellipticine“.....	22
2.2. Vyhledávání pojmu „serine racemase“.....	24
2.3. Vyhledávání pojmu „Lowry method“.....	26
3. Výsledky hledání.....	28
3.1. Pojem „ellipticine“.....	28
3.2. Pojem „serine racemase“.....	33
3.3. Pojem „Lowry method“.....	35
4. Diskuse.....	38
5. Závěr.....	41
6. Slovníček základních pojmů.....	42
7. Seznam použitých zkratk.....	43
8. Seznam použité literatury.....	44
9. Příloha.....	47

Úvod a cíl práce

V minulých letech došlo k obrovskému rozvoji internetu. A stejně jako beletristické knihy, i knihy a časopisy vědecké se postupně začaly objevovat na internetu. Hledání ve specializovaných databázích na internetu usnadní práci z pohledu času, hledat informace na internetu lze téměř kdykoliv a kdekoliv. Hledání na „webu“ má dvě výhody, a sice čas strávený při vyhledávání článků či publikací o určitém problému na internetu, protože je rozdílem hledat články za delší časové období v knihovně či na internetu, dále poté možnost zadávání složitějších dotazů.

Ale i na internetu existuje konkurence. Je známo obrovské množství bibliografických databází přístupných na „webu“. Každá z nich má své výhody i nevýhody. Některé jsou placené, jiné jsou poskytovány zdarma. Liší se také řazením článků, vyhledávací nabídkou, jednoduchostí obsluhy, některé obsahují jen anglicky psané, jiné i studie publikované v dalších jazycích.

Cílem mé práce je usnadnit výběr databáze a určit, kterou ze čtyř zkoumaných databází (SCOPUS, Web of Science, MEDLINE, Google Scholar) použít pro nejefektivnější vyhledávání informací o biochemické problematice.

Při zvolení špatné databáze existují dvě možnosti výsledku – hledanou informaci vůbec nebude možné nalézt, případně bude nutné potřebné články hledat mezi obrovským množstvím studií nepotřebných. Jde zde totiž o snahu maximalizovat relevanci i zachyt článků najednou. Ať už by vědec narazil na jakýkoliv z těchto problémů, je jisté jen to, že hledání potřebných informací bude časově velmi náročné. Výběr správné databáze se v tuto chvíli stává stěžejním úkolem, nicméně je tato volba vždy kompromisem. Žádná z databází totiž neposkytuje kombinaci uživatelsky výborně přijatelného prostředí, stoprocentně relevantních výsledků a zároveň stoprocentního zachytu.

Svou prací bych tedy ráda pomohla při volbě správné databáze a alespoň minimálně tak pomohla usnadnit práci svým kolegům. Chtěla bych určit, kterou databázi využít při řešení biochemických problémů, ulehčit orientaci v rozhraních a ukázat výhody i nevýhody mnou zkoumaných bibliografických databází. Ráda bych také definovala, jaké jsou hlavní rozdíly mezi databázemi placenými a volně přístupnými a zda jsou odlišnosti pouze menší, či jsou mezi těmito dvěma typy rozdíly obrovské.

1. Databáze jako zdroj vědeckých informací

1.1. Srovnávání databází

Bylo publikováno už mnoho studií, které se zabývají srovnáním databází. Jednalo se především o články, které se týkaly jejich porovnání z pohledu určité konkrétní problematiky. Tyto studie se dají rozdělit podle toho, zda se týkají databází faktografických anebo bibliografických. Níže bude popsáno vždy několik již publikovaných článků, chronologicky seřazených.

1.1.1. Faktografické databáze a jejich porovnání

Faktografické databáze jsou takové, jejichž údajové základny tvoří faktické informace.¹⁾ Obsahují konkrétní informace o dané problematice a většinou není nutné dohledávat primární pramen, neboť tyto databáze nabízí přímo hledané informace. Rozdělit je můžeme například podle obsahu²⁾ na:

- numerické databáze – obsahují hlavně statistická data
- faktové databáze – slovně, nebo pomocí tabulek shrnují podstatná fakta z původních pramenů
- databáze typu průvodce – například adresáře firem, katalogy výrobců, rejstříky

Nicméně toto rozdělení není všeobecně přijímáno. Například TDKIV (Česká terminologická databáze knihovnictví a informační vědy) faktografické databáze nijak dále nedělí.¹⁾

Jedním z článků porovnávajícím faktografické databáze je studie³⁾ Biareze a kol. (1991), která byla zaměřena na vyhledávání nežádoucích účinků léků v devíti databázích. Analýza se týkala osmi on-line systémů a jednoho systému off-line – databázi přístupné na CD-ROMu. Jejich práce přinesla zjištění, že schopnost databází poskytovat informace o nežádoucích účincích závisí na kvalitě a velikosti každé databáze zvlášť.

Dalším příkladem může být článek⁴⁾ Guerbeta a Guyoda (2001). Autoři v něm porovnávají vhodnost více než dvou desítek (22) volně přístupných on-line databází pro fyzikálně-chemické, toxikologické a ekotoxikologické vlastnosti chemických látek. Ke srovnání využili následujících databází (zde budou uvedeny zkratky, plné názvy databází jsou uvedeny v seznamu zkratek, kapitola 7): ATSDR, CCRIS, CPDB, ECOTOX, EXTOTOXNET, GAP, GENETOX, HCD, HSDB, HSFS, IARC, ICSC, IPCS-INCHEM, IRIS, NIOSH-IDLH, NIOSH-PG, NIOSH-DOE, NTP, OSHAPEL, SOLVEDB,

TELETOX, TOX-IN. Autoři poté vybrali kritéria, která se jim zdála nejdůležitější pro hodnocení rizik:

- fyzikální kritéria: molekulová hmotnost, teplota rozkladu, hustota, tlak páry, rozpustnost ve vodě, rozdělovací poměr oktanol/voda, Henryho konstanta, biologická rozložitelnost, adsorpční koeficient, poločas rozpadu, biokoncentrační faktor
- toxikologická kritéria: akutní a chronická toxicita pro savce, genotoxicita, karcinogenita, teratogenita, dráždivost, zdravotnická data a data z pracovního lékařství
- ekotoxikologická kritéria: vodní toxicita pro řasy, korýše a ryby, atmosférická toxicita pro hmyz a ptáky a půdní toxicita pro bakterie, měkkýše a rostliny

Dále vybrali 14 chemických látek, které poté v databázích testovali (pojmy skutečně zadávané do databází jsou uvedeny v závorce v případě, že se od českého ekvivalentu výrazně liší):

- pesticidy – atrazin, malathion
- soli kovů – chlorid kadmiový (Cadmium chloride), dibutylchlorid cínčitý (Di-n-butyl tin(IV) dichloride)
- rozpouštědla – dichlormethan (dichloromethane), benzen
- chemicky nasyntetizované sloučeniny – 2-chlorfenol (2-chlorophenol), methylisokyanát (methylisocyanate)
- léky – sulfamethoxazol, bleocin
- chemické rodiny – xyleny (xylenes), rtuť (mercury)
- náhodně vybrané produkty – arprinocid, 2-(2-butoxyethoxy)ethanol

Následně byly databáze testovány pro 14 chemických látek na základě výše uvedených kritérií.

Pro zhodnocení kvality databáze autoři zavedli dva indexy – Qi index (ukazující kvalitu informací) a Pi index (hodnotící „sílu informací“). Indexy počítali následovně:

$$Q_i = \frac{\text{počet nalezených záznamů o látkách v databázi}}{27 \times \text{počet nalezených látek}}$$

$$P_i = \frac{\text{počet nalezených záznamů o látkách v databázi}}{27 \times 14}$$

V těchto vzorcích odpovídá hodnota 27 počtu kritérií, 14 je potom počet hledaných látek.

Příklad: Pokud databáze poskytne 180 záznamů o devíti zkoumaných chemických látkách, lze tyto indexy spočítat takto:

$$Q_i = \frac{180}{27 \times 9} = 0,74; P_i = \frac{180}{27 \times 14} = 0,48$$

Výsledky se poté dají pomocí indexů podle autorů hodnotit následovně:

- nízké Q_i , nízké P_i – databáze, která našla jak malé množství chemických látek, tak malé množství informací o nich
- vysoké Q_i , nízké P_i – specializovaná databáze obsahující velké množství záznamů o malém množství látek
- nízké Q_i , vysoké P_i – databáze obsahuje velké množství hledaných látek, ale malé množství informací o nich
- vysoké Q_i , vysoké P_i – nejlepší databáze, pokrývající velké množství chemických látek, o kterých lze nalézt hodně informací

Výsledkem práce Guerbeta a Guyoda byly dvě obsáhlé tabulky. První srovnává jednotlivé databáze – kolik chemických látek můžeme nalézt, zda obsahují bibliografické údaje, zda je hledání snadné, a také uživatelský komfort („user-friendliness“). Tabulka druhá shrnuje další výsledky výzkumu – uvádí přesný počet kritérií, které jednotlivé databáze našly. Jako nejlepší vyhodnotili autoři databázi HSDB (Hazardous Substances Data Bank), která měla vyrovnané nejvyšší indexy – Q_i 0,71 a P_i 0,62. Jako nejhorší vyšly ze studie TOX-IN, TELETOX a GENETOX.

Tuto studii později autoři rozšířili tak⁵⁾, že srovnávají shodným způsobem 35 databází (jak přístupných on-line na internetu, tak off-line na CD-ROMu); opět z hlediska řešení environmentálních rizik. Výsledkem jejich práce bylo zjištění, že žádná z prohledávaných databází nedala stoprocentní výsledky – pouze několik z databází se ukázalo vhodných pro řešení tématu z daného oboru a ani u chemických látek nebyl výsledek příliš příznivý, žádná z databází totiž neposkytla články ke všem 14 vybraným látkám. Nejlépe byla opět vyhodnocena HSDB.

Dalším typem článků o faktografických databázích jsou studie o jejich vzájemném překryvu. Takové studie říkají, do jaké míry se shoduje záchyt databází. Shrnutí již publikovaných textů nabídli ve svém článku z roku 2003 W.H.Hood a C.S.Wilson⁶⁾:

Hawkins a Miller⁶⁾ (1977) se zaměřili na informace přístupné on-line. Prohledávali 8 databází aby zjistili, které z nich poskytnou přístup ke třiceti předem

vybraným vědeckým časopisům. Výsledkem bylo zjištění, že 67,3% relevantních záznamů lze nalézt pouze v jedné z osmi databází.

Podobný závěr nabídl Ernest a kol.⁶⁾, když roku 1988 publikovali studii, jejímž závěrem bylo, že „zdvojení mezi databázemi je relativně malé“, tedy že pouze malé množství informací je ve více než jedné databázi.

Další z těchto článků o překryvu databází jsou například studie Nichollse⁶⁾ (1989), Walkera⁶⁾ (1990) a Mychko-Megrina⁶⁾ (1991).

V samotné studii Hooda a Wilsona⁶⁾ bylo výsledkem zjištění, že 63% témat je schopna nalézt pouze jedna z testovaných databází, zbylých 37% záznamů má překrytí u 2-12 databází. Pro úplnost dodejme, že autoři prohledávali 20 databází z nejruznějších oborů. Rozmanitost oborů je nejspíše důvodem, proč zadané výrazy našlo maximálně 12 databází.

Zhodnotím-li situaci ohledně faktografických databází, je překryv relativně malý, proto je nutné při hledání využít více než jednu databázi.

1.1.2. Bibliografické databáze a jejich porovnání

Bibliografické databáze většinou vznikaly jako elektronické verze referátových časopisů.²⁾ Jejich datovou základnu tvoří bibliografické informace, vymezené obsahově, typem popisovaných zdrojů nebo jejich lokací. Slouží především k vyhledávání bibliografických informací. Obsah je uložen v jednotně strukturovaných bibliografických záznamech, umožňujících vyhledávání podle hodnoty obsažených položek. Pravidla popisu a jeho podrobnost se mohou v různých databázích lišit.¹⁾ Bibliografické záznamy dokumentů jsou doplněny abstrakty s klíčovými slovy popisujícími obsah dokumentů. Jedná se tedy o sekundární informaci o existenci primárního informačního pramene, jehož plný text v databázi již nemusí být a je třeba jej případně následně získat.²⁾

I o bibliografických databázích bylo publikováno nemalé množství srovnávacích studií, v nichž se ale spíše většina týkala jiného oboru než biochemie.

Porovnáním bibliografických databází pro řešení problémů z oboru psychiatrie se zabývala práce⁷⁾ McDonalda a kol. z roku 1999. Autoři si vybrali jako klíčová slova „psychiatry and neurology“ a ta byla zadávána do bibliografického „balíčku“ (bibliographic software package) ProCite (Uhlrich's International Periodicals directory). Toto zadání bylo pro zjednodušení takové, že nebylo vepsáno celé slovo, ale pouze kořen slova s tzv. rozšířením zprava, tedy „psych*“ nebo „neuro*“, což umožní nalézt

všechny články týkající se slov s těmito kořeny. Program všechny nalezené záznamy poté rozdělil na dvě oblasti – jednu týkající se psychiatrie a druhou neurologie.

Celkem bylo nalezeno 1431 publikací, z čehož bylo 977 z oboru psychiatrie (z toho 180 bylo zařazeno současně do neurologie), 496 neurologie a 138 bylo označeno jako nejasné (jednalo se především o newslettery, případně nebyl k dispozici abstrakt). Tyto články byly převážně ze čtyř bibliografických databází: PsycLIT, EMBASE, Biosis a MEDLINE. Z článků řazených do oboru psychiatrie jich 506 (52%) náleželo do více než jedné z výše uvedených databází. Ve všech databázích bylo nalezeno 122 článků.

Nejvíce záznamů celkem nalezla databáze PsycLIT. Autoři usoudili, že pro úspěšné hledání informací v oborech psychiatrie a neurologie poslouží nejlépe buď tato databáze sama (nalezla 73% z celkového počtu výsledků), nebo ve spojení s EMBASE (dohromady by bylo pokrytí 91%), Biosis (86%), či MEDLINE (85%)

Jedním z novějších článků na toto téma je publikace⁸⁾ Wilkinse a kol. (2005), která srovnává využití databází EMBASE a MEDLINE pro hledání informací z oblasti rodinné medicíny (databázi EMBASE stejně jako MEDLINE vlastní společnost Elsevier a je taktéž databází medicínskou, nabízí údajně o 2000 titulů z oblasti biomedicíny více než databáze MEDLINE a využívá stejný thesaurus, EMTREE⁹⁾). Důvodem tohoto srovnání je fakt, že v severní Americe je lékaři používán převážně MEDLINE, zatímco v Evropě je využíván hlavně EMBASE. Vyhledávány byly v této práci následující pojmy (vyhledávání bylo opět provedeno v angličtině, já zde uvádím český překlad pojmů): diabetes, deprese, zánět středního ucha (otitis media), ischemická choroba srdeční (coronary artery disease), hypertenze, zánět nosohltanu (pharyngitis), ztráta čichu (sinusitis), rýma (rhinitis), bronchitida, astma, infekce močového traktu (urinary tract infection), dermatitida, obezita, osteoartritida, revmatoidní artritida.

Výsledkem práce bylo zjištění, že z více než tři a půl tisíce záznamů (3622) pouze 177 citací sdílí současně MEDLINE i EMBASE, ostatní záznamy jsou pro jednu z databází unikátní. Větší počet jedinečných záznamů má EMBASE (a to 2246 oproti 1199 u MEDLINE). Jako příklad lze uvést hledání pojmu „urinary tract infection“, u něhož našel EMBASE 60 unikátních záznamů, zatímco MEDLINE pouze 25, překryv byl zjištěn pouze u pěti vyhledaných citací. Hledání všech pojmů je na konci citované studie přehledně shrnuto v tabulce, která uvádí, kolik záznamů bylo nalezeno pro jednotlivé pojmy a dané databáze. V závěru práce je ještě uvedena poznámka, že

ačkoliv EMBASE našel více citací než MEDLINE, nevypovídá to nic o kvalitě nalezených záznamů.

Každá z databází je také sama o sobě tématem velkého množství přehledných článků, které nejčastěji obsahují popis každé z nich, její největší výhody a nevýhody, někdy také hodnotí obtížnost a úspěšnost vyhledávání.

1.2. Charakteristika jednotlivých databází

V následujícím textu budou podrobněji popsány jednotlivé databáze, které ve své bakalářské práci budu srovnávat. Informace (historie databází, vlastníci, klasifikace nalezených záznamů apod.) byly vyhledávány jak na internetových stránkách jednotlivých databází, tak v článcích zabývajících se jejich charakteristikou.

1.2.1. SCOPUS



Obr.1: logo databáze SCOPUS

Bibliografická databáze SCOPUS je databází placenou. Je poskytována na webu pro předplatitele a jejím vlastníkem je společnost Elsevier. Poprvé byl SCOPUS otestován v provozu na začátku listopadu roku 2004 na několika univerzitách.¹⁰⁾ Čerpá z okolo 18 000 publikací a časopisů od více než 5 000 autorů z celého světa. Více než polovina příspěvků pochází z Evropy, Latinské Ameriky a Asie a Tichomoří, obnovování záznamů probíhá denně.

Ve SCOPUSu lze nalézt okolo 38 milionů článků, z nichž je 19 milionů z doby před rokem 1996 (některé příspěvky sahají až k roku 1823) a dalších 19 milionů z období od roku 1996 až dodnes¹¹⁾. SCOPUS má vlastní vyhledávací engine, Scirus.

Podle článku Goodmana a Deise z roku 2004¹⁰⁾ leží síla SCOPUSu v přírodovědných oborech. Ve srovnání s Web of Science podle těchto autorů totiž SCOPUS obsahuje daleko více vědeckých publikací v jiném jazyce než angličtině. Bohužel však v jiných oborech (veřejné záležitosti, historie a ostatní humanitní vědy) spíše zaostává.

Záznamy ve SCOPUSu jsou klasifikovány pomocí thesauru EMTREE. Ten obsahuje více než 56 000 termínů (je dvakrát větší než thesaurus MeSH u MEDLINE a současně je MeSH v EMTREE obsažen, viz. Kap. 1.2.3.) a více než 230 000 synonym.

Tato synonyma nám například usnadní hledání v případě, že existuje dva a více pojmů pro tutéž sloučeninu. Příkladem může být hledání látky „paracetamol“ – databáze nalezne jak články, ve kterých se vyskytuje pojem paracetamol, ale stejně tak studie, ve kterých je paracetamol nazván acetaminophen. EMTREE obsahuje jak názvy chemické, tak obchodní, případně laboratorní a hledání tudíž není zpomalené zadáváním synonym do vyhledávacích polí.¹²⁾

Hledání v databázi SCOPUS není nijak obtížné. Hlavní stránka databáze odkáže čtenáře na základní hledání („Basic search“), kde je nutné zadat vyhledávaný výraz, zvolit, ve které části abstraktu či obsahu článku má být výraz použit (např. authors, article title, abstract apod.) a zadat limity. Těmito limity se rozumí časové ohraničení a obor, ve kterém má být pojem vyhledán (life sciences, health sciences, physical sciences, social sciences and humanities). Pokud by tato ohraničení zadána nebyla, automaticky by byl výraz prohledán ve SCOPUSu ve všech letech a všech čtyřech oborech.

The screenshot displays the SCOPUS Basic Search interface. At the top, the SCOPUS logo is visible, followed by navigation tabs: Search, Sources, Analytics, My Alerts, My List, and My Settings. A banner promotes the Design Partner Program. The main search area includes tabs for Basic Search, Author Search, Affiliation Search, and Advanced Search. The Basic Search tab is active, showing a search bar with the text 'ellipticine' and a dropdown menu set to 'Article Title, Abstract, Keywords'. Below the search bar, there is a 'Limit to:' section with three main categories: Date Range (inclusive), Document Type, and Subject Areas. The Date Range section has two radio buttons: 'Published' (selected) and 'Added to Scopus in the last 7 days'. The Document Type section has a dropdown menu set to 'All'. The Subject Areas section has four checkboxes, all of which are checked: 'Life Sciences (> 4,300 titles)', 'Health Sciences (> 6,800 titles) Includes 100% Medline coverage', 'Physical Sciences (> 7,200 titles)', and 'Social Sciences & Humanities (> 5,300 titles)'. A 'Search' button is located at the bottom right of the search area. At the bottom of the interface, there is a 'Search History' section with a 'Combine' button and a 'Go' button.

Obr.2: základní možnosti vyhledávání v databázi SCOPUS

Kromě základního hledání existují ještě další možnosti vyhledávání:

- Vyhledávání autora („Author Search“) - můžeme hledat nejen podle jména a příjmení autora, ale také podle místa jeho působení, tzn. můžeme do pole

„Affiliation“ zadat například název univerzity, na které hledaný člověk působí či působil

- Vyhledávání instituce („Affiliation search“) – využívá se v případech, kdy chceme najít určitého vědce a víme jen to, že působí či působil na určité univerzitě
- Rozšířené vyhledávání („Advanced search“) – vhodné pro pokročilého uživatele bibliografických databází, zde již nejsou možnosti zadání jako v hledání základním, v této části zadáváme již úplné výrazy včetně Booleových operátorů a dalších pomocných výrazů („Codes“ – jedná se o kódy typů databázových polí), výhodou databáze SCOPUS je fakt, že po kliknutí na jakýkoliv z kódů je nabídnuto vysvětlení

SCOPUS se od ostatních databází odlišuje několika vlastnostmi – například i ve vyhledaných souborech lze aplikovat omezení – i po zadání výrazu tedy lze ohraničit hledání na určitý rok, autora, či časopisu, anebo na všechny tyto podmínky dohromady, aniž by bylo nutné se vrátit na původní stránku. PubMed, ani Google Scholar nemají také velmi užitečnou vlastnost – možnost označit si články, které lze poté nalézt v určité složce. U databáze SCOPUS se takto označené záznamy shromažďují v záložce „My list“.

1.2.2. Web of Knowledge



Obr.3 : logo databáze Web of Knowledge

Web of Knowledge je také placená bibliografická databáze, pokytuje ji Ústav vědeckých informací společnosti Thompson Reuters. Web of Knowledge slučuje obsah tří databází (Web of Science, Journal Citation Reports a Current Contents Connect).¹³⁾

Web of Science

Databáze Web of Science je sestavena ze sedmi menších částí (pro chemika jsou podstatné Science Citation Index, Index Chemicus a Current Chemical Reactions). Co se týče přírodních a technických věd, zdrojem pro tuto databázi je více než 5600 periodik, vyhledávat zadané dotazy lze zpětně až do roku 1900 a aktualizace probíhá jednou týdně s přírůstkem 19 000 nových záznamů.¹⁴⁾

Rozhraní Web of Science bylo největším zdrojem bibliografických údajů až do roku 2004, kdy začal fungovat SCOPUS (téměř ve stejnou dobu začal fungovat i Google Scholar, jenž má oproti Web of Science a SCOPUSu tu výhodu, že je neplacený).¹⁵⁾

Vyhledávání ve Web of Science je stejně jako u ostatních porovnávaných databází jednoduché.

The screenshot shows the Web of Science search interface. At the top, there are navigation links: Search, Cited Reference Search, Structure Search, Advanced Search, Search History, and Marked List (0). Below this is a yellow banner with the text "Web of Science® - with Conference Proceedings". The main search area has a "Search for:" label and three input fields. The first field has a dropdown menu set to "Topic" and an example "oil spill* mediterranean". The second field has a dropdown menu set to "Author" and an example "O'Brian C* OR OBrian C*". The third field has a dropdown menu set to "Publication Name" and an example "Cancer* OR Journal of Cancer Research and Clinical Oncology". There are "AND" buttons between the fields. Below the search fields are "Search" and "Clear" buttons. At the bottom, there are "Current Limits: [Hide Limits and Settings] (To save these permanently, sign in or register.)" and a "Timespan:" section with a radio button for "All Years" (updated 2010-02-27) and a "From" section with a date range "1945-1954" to "2010" (default is all years). Below this are "Citation Databases:" and "Chemical Databases:" sections with checkboxes for various databases.

Obr.4: Úvodní stránka Web of Science

Výhodou této databáze je jednodušší vyhledávání víceslovných výrazů. Zde je totiž na hlavní stránce možné navolit libovolný počet kolonek (například podle počtu slov) a vlevo zvolit jeden z operátorů přímo ve výběru (AND, OR, NOT). Není tedy nutné sestavovat vyhledávací výraz, je sestaven rozhraním. V pravé části lze současně vybrat, v jakém datovém poli se má výraz objevit (Topic, Title, Author, Editor, atd.).

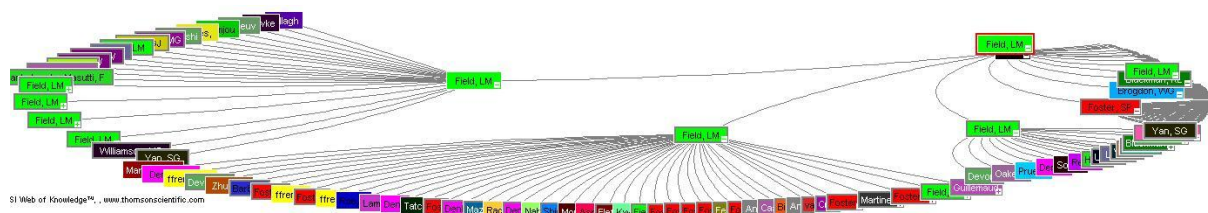
Stejně jako u databáze SCOPUS existuje možnost vybrat hraniční hodnoty – je zde na výběr omezení časové a dále lze zvolit, v jakých databázích se má vyhledávání provádět (jak je popsáno výše, pro chemika jsou nejdůležitější Science Citation Index, Index Chemicus a Current Chemical Reactions).

Popis uvedený výše se vztahuje na základní vyhledávání („basic search“). Web of Science nabízí i další možnost:

- Hledání citovaných referencí („Cited Reference Search“) – lze zadat autora, jméno práce či rok jejího vydání

- Hledání struktury („Structure Search“) – slouží k vyhledání struktury látky, nebo po zakreslení struktury umožňuje zjistit, o jakou molekulu se jedná a najít články s ní spojené
- Pokročilé vyhledávání („Advanced Search“) – v této části je potřeba již určitá zručnost badatele, pro základního uživatele se příliš nehodí. Pojem je zadáván pomocí Booleových operátorů, oddělován závorkami atd., stejně jako u databáze SCOPUS jsou jednotlivé operátory a další zkratky („Field tags“) vysvětleny
- Historie hledání („Search history“) – jednoduše lze nalézt poslední hledané pojmy
- Seznam označených článků („Marked list“) – vyhledané články, které jsou považovány za podstatné, lze označit a převést právě do tohoto seznamu – po ukončení vyhledávání a kliknutí na seznam je zde možné najít důležité nalezené články

Web of Science má několik vlastností, které ho odlišují od ostatních databází. Jednou z nich je již zmíněný seznam označených článků. Tuto vlastnost má ze mnou zkoumaných databází pouze SCOPUS a Web of Science. Stejně jako v databázi SCOPUS ihned při prvním zobrazení výsledků – kde jsou vidět pouze názvy článků – víme, zda a kolikrát byl daný záznam citován. Web of Science také umožňuje vytvořit citační mapu (funkce „Citation map“, je k nalezení po rozkliknutí abstraktu článku), která graficky znázorňuje, kým byl daný článek již citován. Příklad citační mapy je uveden níže, na Obr.5.



Obr.5: dvougenerační citační mapa z Web of Science k článku:

FIELD, L.M.; CRICK, S.E.; DEVONSHIRE, A.L. *Web of Knowledge* [online]. 1996 [cit. 2010-02-27]. Polymerase chain reaction-based identification of insecticide resistance genes and DNA methylation in the aphid *Myzus persicae* (Sulzer)

1.2.3. Rozhraní PubMed databáze MEDLINE



PubMed má neplacený přístup a je rozhraním k databázi MEDLINE. Obsahuje abstrakty a často i úplné články

Obr.6: logo rozhraní PubMed z biomedicínského oboru, jak v angličtině, tak v národních jazycích (u cizojazyčných článků je název v hranatých závorkách). PubMed je provozován United States National Library Medicine (NLM) při National Institutes of Health (NIH). Články jsou zde doplňovány dvakrát týdně – vždy v úterý a sobotu.¹⁶⁾

Protože je PubMed rozhraním k databázi MEDLINE, nejprve bych se několika větami ráda zmínila o ní. Databáze MEDLINE je jedním z nejdůležitějších zdrojů informací týkajících se lékařství a dalších vědních oborů, jako je biologie či chemie. Obsahuje 13 miliónů záznamů, a to od roku 1966 až do současnosti. Jde především o citace, abstrakty celosvětové lékařské literatury včetně výzkumu, klinické praxe, administrativy a služeb pro ochranu zdraví.¹⁷⁾

Záznamy v MEDLINE jsou klasifikovány pomocí thesauru MeSH – Medical Subject Headings. Skládá se ze sady přiřazených výrazů (zvaných deskriptory) v hierarchické struktuře, která umožňuje prohledávání na různých stupních. Kromě hierarchického přístupu jsou MeSH deskriptory současně uspořádány v abecedním pořadí. Na nejobecnější úrovni jsou velmi používané a široké pojmy, jako například „anatomie“, či „mentální poruchy“. Konkrétnější deskriptory mohou být například „kotník“, či „porucha chování“, a jsou až na konci jedenáctistupňové hierarchie. V roce 2009 existovalo v MEDLINE 25 186 deskriptorů.¹⁸⁾

PubMed je rozhraní, které je neplacené – a to díky tomu, že je financován americkou vládou. Prezident George Bush ml. totiž podepsal v prosinci roku 2007 zákon „Consolidated Appropriations Act“, který obsahuje ustanovení, že NIH (National Institutes of Health) musí poskytovat veřejnosti on-line přístup k výsledkům výzkumu financovaného z veřejných zdrojů - americká vláda tím povolila přístup k výzkumu veřejnosti¹⁹⁾. Tento zákon později podepsal i prezident Obama (v březnu roku 2009), čímž je zaručeno, že fulltexty v PubMedu (od článků financovaných právě z veřejných peněz) jsou nyní volně přístupné. Zákon z roku 2009 mimo jiné také ukládá, že se na PubMedu objeví plné texty článků nejpozději 12 měsíců po jejich zveřejnění v tištěném časopise.²⁰⁾

Vyhledávání v PubMedu není o mnoho obtížnější než v jakékoliv jiné databázi. Po otevření úvodní stránky se zobrazí pouze omezená nabídka – do vyhledávacího pole

lze zadat daný výraz a současně zadat, zda chceme prohledat PubMed, databázi proteinů, nukleotidů, atd. Po rozkliknutí nabídky „Limits“ se zobrazí různé možnosti, kterými můžeme zpřesnit vyhledávání. Jelikož je PubMed databáze, resp. rozhraní medicínské, obsahuje množství hraničních možností, které chemické databáze neumožňují. Například lze zadat, zda jde o studie na zvířatech či lidech, nebo také specifikovat pohlaví zkoumaných subjektů. Dále lze omezit hledání podobně jako v ostatních databázích – datum a typ článku (review, klinické studie, ...), či jazykem, ve kterém je článek napsán (nicméně abstrakty jsou v angličtině). Můžeme také omezit hledání na výsledky, které obsahují kromě abstraktů i plné texty.

Co se týče rozšířeného vyhledávání („Advanced Search“), platí pro něj totéž jako u předchozích databází: je zde již potřebná jistá zručnost zadavatele, ale ze všech čtyř databází je toto rozšířené vyhledávání dle mého názoru nejjednodušší – zadá se výraz, část článku, ve kterém se pojem má nacházet (title, abstract, author, ...) a dále Booleových operátorů AND, OR nebo NOT a tento postup lze opakovat nesčetněkrát, dokud dotaz není vytvořen.

Jistou výhodou v rozšířeném hledání je volba „index“ – například po zadání vyhledávacího výrazu „Crick“ a zvolení místa výskytu „Author“ se nám po kliknutí na tuto volbu „Index“ zobrazí seznam všech autorů, jejichž příjmení je Crick a v PubMedu se vyskytuje alespoň jeden jejich článek.

Zatímco u databází Web of Science a SCOPUS existuje schránka, do které si mohu přesunout všechny zajímavé články, u PubMedu přímo takováto volba není, ale vybrané články lze například zaslat na e-mail či je uložit. Je poté již na samotném uživateli, která možnost je pro něj výhodnější a využitelnější.



1.2.4. Google Scholar

Databáze Google Scholar patří, jak je už z názvu patrné, společnosti Google. V současné době je na internetu přístupná zkušební beta verze, a to od listopadu r. 2004.²¹⁾ Google Scholar není placený,

Obr.7: logo databáze Google Scholar jeho reklamním sloganem je „Stand on the Shoulders of giants“, tedy „Stůjte na ramenou obrů“, což je dle autorů „uznání toho, že vědecký výzkum většinou staví na tom, co již bylo objeveno“ a pochází z citátu Sira Isaaca Newtona – „Viděl jsem daleko, protože jsem stál na ramenou obrů.“²²⁾

Google Scholar nabízí vyhledávání nejen v odborných člancích, ale také například v disertačních pracích umístěných na internetu, v knihách, nebo na stránkách, které jsou dle autorů považovány za vědecké. Na rozdíl od ostatních databází, u Google Scholar jde o fulltextové vyhledávání, což zajišťuje velký počet vyhledaných záznamů (toto dokazuje nejen jeden článek²³⁾).

Prvním podstatným problémem je, že Google Scholar prochází při vyhledávání jen časopisy, u nichž vydavatelé prohledávání povolí. Například vydavatelství Elsevier, které je poskytovatelem databáze SCOPUS, umožnil prohledávání některých svých časopisů až od roku 2007. Roku 2008 ještě nebyly pro Google Scholar zpřístupněny časopisy společnosti American Chemical Society.²⁴⁾ Jak je tomu nyní se lze jen domýšlet – nikde není zveřejněn seznam časopisů či internetových stránek, které Google Scholar prohledává, a stejně tak není známá frekvence obnovování článků, což je samozřejmě další nevýhoda databáze. V souvislosti s tímto zároveň narážíme na další problém - jelikož Google Scholar vyhledává i v placených časopisech, po kliknutí na nalezený text se objeví často jen abstrakt, či například pouhý první odstavec článku.

Pokud jde o vyhledávání a řazení příslušných článků, autoři Google Scholar o tomto tématu tvrdí následující: „Cílem služby Google Scholar je třídít články tak, jak to dělají vědečtí pracovníci – vzít v potaz celý text každého článku, autora, publikační zdroj, ve kterém se článek objevil a kolikrát byl citován v jiné vědecké literatuře. Nejrelevantnější výsledky se vždy objeví na první stránce.“⁴⁾

Nezanedbatelnou výhodou databáze je pro nás možnost vyhledávání v češtině. Níže jsou uvedeny úvodní stránky Google Scholar jak v češtině, tak v angličtině, dále se ovšem budu věnovat anglické verzi databáze.



Obr.8: Výchozí stránka Google Scholar
česká verze



Obr.9: Výchozí stránka Google Scholar
anglická verze

Jak je vidět již při porovnání úvodních stránek (obrázky 8 a 9), anglická verze nabízí hned v „basic search“ výběr, zda si člověk přeje vyhledat zadaný pojem ve vědních článcích, či v právnické literatuře.

Podstatnější a lépe využitelné je vyhledávání rozšířené – „Advanced search“.

Obr 10: rozšířené vyhledávání v Google Scholar

Pokud existuje víceslovný název, který je nutné vyhledat, lze si vybrat mezi čtyřmi variantami vyhledávání. Záleží na tom, zda chceme, aby vyhledaný článek našel celý víceslovný název, či alespoň jedno slovo z hledaného výrazu a kde ono slovo má být použito (kdekoliv v článku, či v názvu článku). Google Scholar také obsahuje možnost zadat pojem a určit, že jej nesmí vyhledaný článek obsahovat.

I Google Scholar obsahuje možnosti omezení, které lze zvolit, i v tomto případě se jedná o rok publikace a obor, ze kterého má hledaný záznam být. Hledání lze také omezit, stejně jako u ostatních databází, zadáním jména autora či časopisu.

Výhody, které na bibliografické databázi Google Scholar prozatím shledávám, je přítomnost nápovědy téměř v každém kroku vyhledávání, rychlost databáze a jednoduchost zadávání vyhledávacího dotazu.

2. Metody vyhledávání

Pro porovnání databází byly zvoleny modelové problémy. Jeden pojem byl zvolen velmi jednoduchý, jedná se o jednoslovný název látky – „ellipticine“ – a to z toho důvodu, že se studií této sloučeniny zabývá jedna z vědeckých skupin na naší fakultě. Další dva problémy byly složitější, neboť se jednalo o dvouslovné výrazy. Jednalo se o dvouslovný biochemický název sloučeniny „serine racemase“, který byl vybrán z důvodu, že se do literatury dostal až v poslední době a bylo zajímavé pozorovat, kolik relevantních článků se v daném roce (2008) objevilo. Nejobtížnější problém byl „Lowry method“, kdy bylo zadání stížené faktem, že existuje mnoho autorů jménem „Lowry“, stejně tak vyhodnocení bylo obtížnější než u předchozího pojmu, neboť počet výsledků byl očekáván výrazně vyšší než u předchozích modelových problémů.

2.1. Vyhledávání pojmu „ ellipticine“

První byl vyhledáván jednoduchý typový dotaz, konkrétně bylo zvoleno protinádorové agens ellipticin. Hledání bylo omezeno pouze rokem publikování, kterým byl rok 2008. Výsledky jsou ke dni 13.3.2010.

SCOPUS

Prvním prohledávanou databází byl SCOPUS společnosti Elsevier. Pojem byl zadán v základním hledání, místo výskytu bylo zvoleno „Article Title, Abstract, Keywords“. Limitem byl zvolen jednak rok 2008, jednak jsem hledání omezila na „Life Sciences“, „Health Sciences“ a „Physical Sciences“, možnost „Social Sciences & Humanities“ byla zanedbána.

Přesný vyhledávací dotaz:

```
TITLE-ABS-KEY(ellipticine) AND SUBJAREA(mult OR agri  
OR bioc OR immu OR neur OR phar OR mult OR medi OR  
nurs OR vete OR dent OR heal OR mult OR ceng OR CHEM  
OR comp OR eart OR ener OR engi OR envi OR mate OR  
math OR phys) AND PUBYEAR IS 2008
```

Takovéto zadání poskytlo v databázi SCOPUS 36 výsledků, nejvíce článků se týkalo farmakologie a toxikologie (18) a biochemie, genetiky a molekulární biologie (16 výsledků).

Web of Science

V rozhraní Web of Science nebylo vyhledávání o mnoho náročnější než u databáze předchozí. Dotaz byl zadán do vyhledávacího prostoru, místo výskytu daného pojmu bylo zvoleno „Topic“. Kromě omezení rokem 2008 byly limitem prohledávané databáze – v tomto případě byly zvoleny „Science Citation Index“, „Index Chemicus“ a „Current Chemical Reactions“.

Přesný vyhledávací dotaz:

```
Topic=(ellipticine)
Timespan=2008. Databases=SCI-EXPANDED, CPCI-S, IC,
CCR-EXPANDED.
```

Po zvolení všech těchto limitů bylo výsledkem hledání 30 záznamů, z čehož se 11 týkalo organické chemie, a v porovnání s databází předchozí jich pouze 5 bylo z oboru farmakologie a farmacie. Vyhledané záznamy se vyskytly ve třech kategoriích: 28 z 30 výsledků byly články zařazené do kategorie „article“, 1 „review“ a 1 „proceeding paper“.

PubMed

V rozhraní PubMed je, v případě že je třeba zadat omezení, zvolena nabídka „Limits“. Zde jsem zvolila časové rozmezí (v nabídce „Dates“ zadáno „Specify date range“ a objeví se možnost zadat období od-do) 1.1.2008 – 31.12.2008. Jiné omezení zadáno nebylo.

Přesný vyhledávací dotaz:

```
("ellipticine"[Substance Name] OR "ellipticine"[All
Fields]) AND ("2008/01/01"[PDAT] : "2008/12/31"[PDAT])
```

Výsledkem hledání bylo 18 nalezených článků souvisejících s ellipticinem. PubMed bohužel nenabízí srovnání, z jakého oboru jsou vyhledané články.

Google Scholar

U vyhledávání pojmu „ellipticine“ v Googlu Scholar jsem využila rozšířené vyhledávání. Místo výskytu hledaného slova jsem zvolila „Anywhere in the article“ a rok byl, stejně jako u předchozích hledání, 2008. Nastavila jsem hledání na následující vědní obory: „Biology, Life Sciences and Environmental Science“, „Chemistry and Materials Science“ a „Medicine, Pharmacology and Veterinary Science“. Ostatní obory jsem zhodnotila jako nepřínosné pro toho hledání (jednalo se o „Business, Administration, Finance and Economics“, „Engineering, Computer Science and Mathematics“, „Physics, Astronomy and Planetary Science“ a „Social Sciences, Arts and Humanities“). Po ukončení vyhledávání jsem ještě nechala zvolenou možnost „include citations“, která zajišťuje, že budou hledány i články, které obsahují citace na studie, ve kterých se vyskytuje hledané slovo.

Přesný vyhledávací dotaz bohužel není možné uvést, v databázi jej totiž nelze nalézt.

Výledkem tohoto zadání bylo 207 nalezených odkazů. Porovnání nalezených výsledků pro Google Scholar a ostatní databáze je uvedeno v následující části.

2.2. Vyhledávání pojmu „serine racemase“

Dále byl do rozhraní databází zadáván pojem „serine racemase“, který by měl být příkladem obtížnějšího pojmu – jedná se o dvouslovný název, který není úplně jednoznačný. Časovým ohraničením byl opět rok 2008, výsledky hledání jsou ke dni 10.4.2010.

SCOPUS

Zadávání tohoto pojmu do rozhraní databáze SCOPUS se shodovalo se zadáváním pojmu „ellipticine“. „Serine racemase“ bylo zadáno do vyhledávacího pole, místo výskytu, jako vždy, „Article Title, Abstract, Keywords“, omezen rok a prohledávané okruhy.

Přesný vyhledávací dotaz:

```
TITLE-ABS-KEY(serine racemase) AND SUBJAREA(mult OR  
agri OR bioc OR immu OR neur OR phar OR mult OR medi  
OR nurs OR vete OR dent OR heal OR mult OR ceng OR  
CHEM OR comp OR eart OR ener OR engi OR envi OR mate  
OR math OR phys) AND PUBYEAR IS 2008
```

Výsledkem hledání bylo 22 záznamů.

Web of Science

Ani ve Web of Science se zadání nelišilo. I zde bylo zadávání pojmu do rozhraní totožné jako u hledání pojmu „ellipticine“.

Přesný vyhledávací dotaz:

```
„Topic=(serine racemase)  
Timespan=2008. Databases=SCI-EXPANDED, CPCI-S, IC,  
CCR-EXPANDED.“
```

Vyhledáno bylo 33 záznamů.

PubMed

I v rozhraní PubMed se zadávání od prvního hledání nelišilo. Naprosto stejně bylo specifikováno časové období.

Přesný vyhledávací dotaz:

```
("serine racemase"[Substance Name] OR "serine  
racemase"[All Fields]) AND ("2008/01/01"[PDAT] :  
"2008/12/31"[PDAT])
```

Databáze MEDLINE poskytla 18 výsledků.

Google Scholar

Ani u databáze společnosti Google se zadání nelišilo. V rozšířeném vyhledávání byl omezen rok a prohledávané okruhy.

Databáze Google Scholar poskytla 289 výsledků.

2.3. Vyhledávání pojmu „Lowry method“

Jako třetí hledaný pojem jsem vybrala hojně využívanou Lowryho metodu pro stanovení proteinů. Pro tento pojem jsme se rozhodli proto, že by výsledkem neměl být úzký okruh záznamů jako u hledání předchozích a samotný dotaz je pro zadání obtížnější než již zmíněné ellipticin, či serinová racemasa. Vyhledávání je také ztíženo faktem, že pojem „Lowry“ nemusí značit jen metodu, ale existuje několik autorů s tímto jménem a tuto volbu je tedy třeba vždy vyloučit. Vyhledávání bylo opět omezeno na rok 2008 a prohledány byly dílčí databáze a obory stejné jako u předchozích hledání obou pojmů (tzn. vynechání Arts & Humanities apod.). Výsledky jsou ke dni 5.4.2010.

SCOPUS

V databázi SCOPUS je vyloučení jména autora velmi jednoduché. Do prvního vyhledávacího pole je zadán pojem „Lowry method“ a místo určení „Article Title, Abstract, Keywords“, tedy zadání naprosto stejné jako v předchozích případech. Pro vyloučení autora je přidáno druhé vyhledávací pole, do kterého je vepsáno jméno „Lowry“, místo určení „Authors“, ale je zvolen Booleův operátor „AND NOT“. Tímto se jednoduše zaručí, že vyhledané záznamy budou opravdu pro Lowryho metodu a zároveň se nebudou vyskytovat články jakéhokoliv autora jménem Lowry.

Přesný vyhledávací dotaz:

```
(TITLE-ABS-KEY(lowry method) AND NOT AUTHOR-  
NAME(lowry)) AND SUBJAREA(mult OR agri OR bioc OR immu  
OR neur OR phar OR mult OR medi OR nurs OR vete OR  
dent OR heal OR mult OR ceng OR CHEM OR comp OR eart  
OR ener OR engi OR envi OR mate OR math OR phys) AND  
PUBYEAR IS 2008
```

Toto zadání poskytlo 36 výsledků, převážně z medicíny (15) a biochemie, genetiky a molekulární biologie (11).

Web of Science

Ve Web of Science je vyloučení autora podobně jednoduché. Do prvního vyhledávacího pole je opět zadáno „Lowry method“, místo výskytu „Topic“. Do druhého pole je vyplněno pouze „Lowry“, místo určení „Author“ a opět zvolen Booleův operátor „NOT“.

Přesný vyhledávací dotaz:

```
Topic=(Lowry method) NOT Author=(Lowry)
Timespan=2008. Databases=SCI-EXPANDED, CPCI-S, IC,
CCR-EXPANDED.
```

Takto zadáný výraz našel v databázi 25 záznamů, převážně ze skupiny biochemických metod výzkumu (5 článků). Z daných textů se jednalo ve většině (24) o „article“, jeden záznam byl označen jako „proceeding paper“.

PubMed

U rozhraní PubMed vyloučení některého z pojmů vypadá obtížněji, ale není o mnoho náročnější než u předchozích zadání. Je nutné využít rozšířené vyhledávání. Nejdříve je do pole vepsán pojem „Lowry method“ (hledáno v „All fields“) a zvolen Booleův operátor „AND“. Poté zadáno místo určení „Author“, opět vepsáno „Lowry“ a zvolen operátor „NOT“.

Přesný vyhledávací dotaz:

```
(Lowry[All Fields] AND ("methods"[MeSH Terms] OR
"methods"[All Fields] OR "method"[All Fields])) NOT
Lowry[Author] AND ("2008/01/01"[PDAT] :
"2008/12/31"[PDAT])
```

Po omezení časového období je výsledkem hledání 25 záznamů. Jak již bylo řečeno při hledání předchozího pojmu „ellipticine“, PubMed neukazuje rozdělení článků do kategorií.

Google Scholar

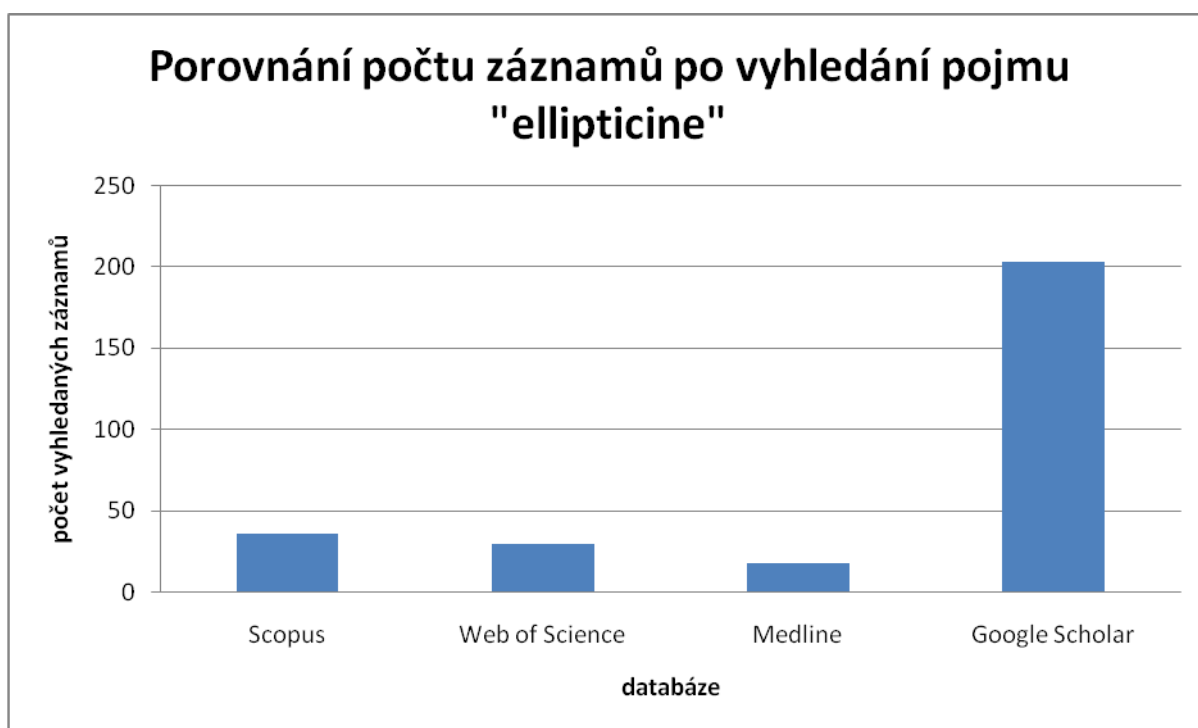
I v Google Scholar existuje volba, pomocí níž můžeme vyloučit některé ze slov dotazu. Problémem však je, že u této databáze nelze vyloučit slovo „Lowry“, u něhož je místo výskytu „Author“. Zde existuje pouze varianta, že vyloučíme nějaké slovo úplně. Při tomto hledání jsem tedy využila kolonku, u které je popis „with the exact phrase“. Tím by mělo být jméno autora vyloučeno. Místo výskytu je, stejně jako v hledání minulých, „anywhere in the article“.

Databáze Google Scholar poskytla tímto zadáním (samozřejmě ještě po omezení časového období a vyloučení několika oborů) 1360 výsledků.

3. Výsledky hledání

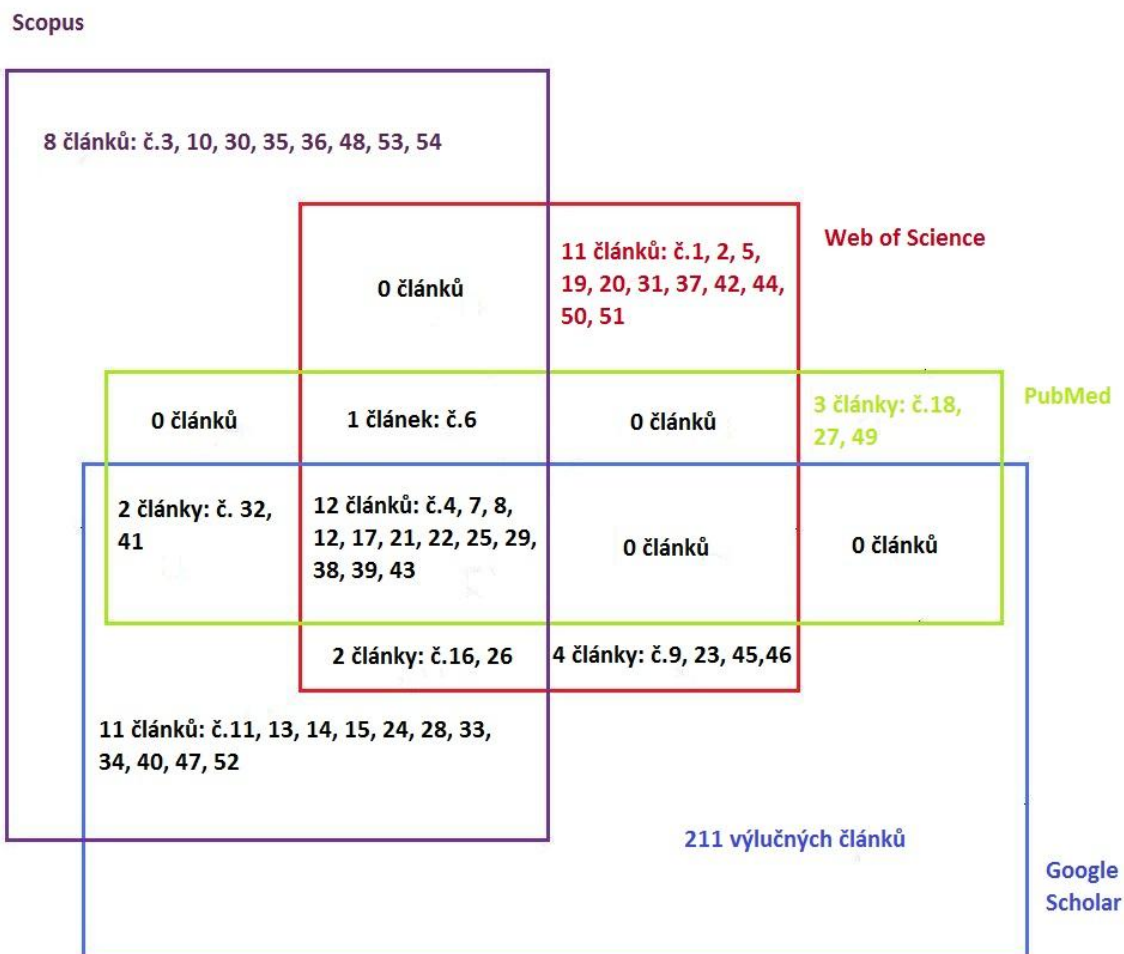
3.1. Pojem „ellipticine“

Jak je vidět z následujícího grafu, ve kterém je porovnán počet vyhledaných záznamů pro výraz „ellipticine“, byla z hlediska počtu nalezených článků nejúčinnější databáze Google Scholar.



Obr.11: Porovnání počtu vyhledaných záznamů pro pojem „ellipticine“

V příloze této práce lze dohledat v abecedním pořadí články nalezené v jednotlivých databázích. Každému článku je přiřazeno číslo a dle těchto čísel lze poté dále články identifikovat v následujícím obrázku 12, kde je ukázán překryv jednotlivých databází.



Obr.12 – zobrazení překryvu článků v jednotlivých databázích

Z obrázku 12 lze vyčíst několik informací. Tou zásadní je, že pouze 12 článků bylo nalezeno všemi čtyřmi databázemi. Po přečtení abstraktů těchto dvanácti článků jsem usoudila, že deset z nich je plně relevantních a ellipticinu se týká. Jedna ze studií (č.43 viz příloha) se však zabývá pouze syntézou a aktivitou protinádorových látek, které byly nově syntetizovány. Ellipticin zde není přímo tématem článku, byl pouze referenční látkou při testování cytostatických účinků nových molekul. Dalším článkem, vyhodnoceným jako nerelevantní, byla studie č.22 (viz příloha), ellipticin je zde využit pouze jako inhibitor.

Nejvíce výlučných článků bylo nalezeno databází Google Scholar (176). Druhou databází u tohoto srovnání byla databáze Web of Science s jedenácti studiemi, které

nebyly jinde nalezeny. Třetí byl SCOPUS s osmi záznamy, a nejméně originálních článků bylo nalezeno v databázi MEDLINE.

Dále se budu podrobněji zabývat právě těmito originálními články, které byly nalezeny pouze jednou z databází.

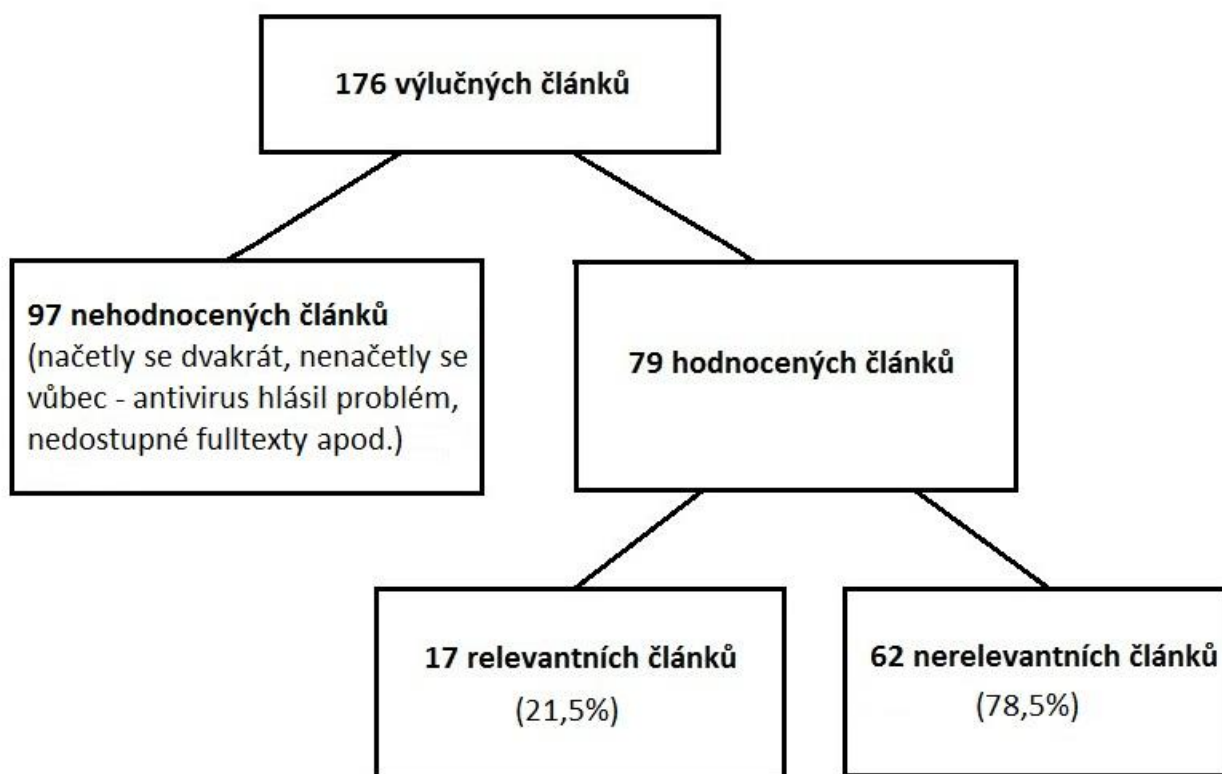
Jak již bylo řečeno, v databázi SCOPUS bylo k roku 2008 nalezeno osm článků přítomných pouze v této databázi společnosti Elsevier. V žádném z abstraktů k těmto článkům nebylo k nalezení slovo „ellipticine“. Tyto záznamy byly vyhledány proto, že pojem „ellipticine“ se v každém z nich objevil v klíčových slovech, tedy „Index keywords“, v kategorii „EMTREE drug terms“ (více o thesauru EMTREE, viz Kap. 1.2.1.). V pěti případech nebylo možné určit relevanci článků, neboť byl nedostupný plný článek. Ze tří dostupných plných textů byl jeden článek vyhodnocen jako relevantní (č.36 viz příloha), další dva jako nerelevantní (č.53 a 54 viz příloha). Unikátní články v databázi Web of Science také neobsahovaly pojem „ellipticine“ v abstraktu, záznamy byly též nalezeny díky přítomnosti tohoto pojmu v klíčových slovech (zde nazvána „Keywords Plus“). V případě, že byl celý článek dostupný, hledala jsem místo výskytu pojmu „ellipticine“. Pokud tento pojem ve studii byl nalezen, jednalo se o naprosto bezvýznamný výskyt – například v článku č.19 (viz příloha) je na ellipticin odkazováno pouze proto, že v jiné práci stejné skupiny byla připravena zmíněná látka právě z ellipticinu. V několika člancích – například č.5 či 20 (viz příloha) se v plném článku dokonce ellipticin nevyskytuje vůbec, je zmíněn pouze v klíčových pojmech (Keyword terms).

V rozhraní PubMed databáze MEDLINE byly nalezeny pouze tři originální články, kterým byla přiřazena číselná označení 18, 27 a 49. Ráda bych rozebrala každý z článků zvlášť, neboť jsou v nich značné rozdíly:

- Článek č.18 (viz příloha) – pojednává o apoptóze indukované ellipticinem, pojem se vyskytuje v abstraktu i v MeSH pojmech, k dispozici však nebyl celý článek. Z přístupného obsahu se ale lze domnívat, že studie relevantní je.
- Článek č.27 (viz příloha) – ellipticine není zmíněn v abstraktu, objevuje se pouze v MeSH označeních. To může tedy vést k mylné domněnce, že se článek ellipticinu vůbec netýká. Avšak po pročtení článku lze zjistit, že se ellipticinu týká celá kapitola článku zvaná „vazba ellipticinu na poly(A)“.
- Článek č.49 – u této studie se ellipticin vyskytuje v abstraktu, avšak pojem již není přítomen v MeSH deskriptorech. Plný text bohužel není dostupný, lze nalézt pouze

názvy kapitol, ale i z nich je patrné, že se článek minimálně v jedné kapitole („ryby vystavené působení β NF léčené ellipticinem“) protinádorového léčiva týká.

Poslední analyzovanou databází byl Google Scholar. Zde bylo nalezeno 176 výlučných článků, jež se v ostatních databázích nevyskytovaly. Otázkou tedy zůstalo, zda jsou tyto záznamy relevantní. 12 z těchto článků nešlo načíst (antivirový program hlásil nebezpečí při otevírání), 4 studie se objevily ve vyhledaných záznamech dvakrát. U 81 nalezených odkazů se pojem ellipticine nevyskytoval v abstraktu, ani v klíčových slovech, avšak nebyl k dispozici plný text. Nebylo tedy možné objektivně zhodnotit, zda se jedná o články relevantní, či nerelevantní. 65% článků tedy vůbec nebylo možné analyzovat, vyhodnocovala jsem tedy zbylých 35%, 79 záznamů.



Obr.13: rozdělení unikátních článků u pojmu „ellipticine“ v Google Scholar

Jak je jasné patrné z obrázku č.13, pouze 21,5% z analyzovaných článků bylo relevantních, to znamená, že o ellipticinu se ve studii či kapitole knihy blíže psalo. Zbylých téměř 80% vyhledaných záznamů relevantních nebylo, to znamená, že pojem nebyl nalezen ani v plném textu a když ano, o protinádorovém agens se text zmiňoval jen okrajově. Tyto články byly vyhledány převážně z toho důvodu, že byl například

ellipticin pozitivní kontrolou ve studii, nebo, a to byl případ nejčastější, byl pojem „ellipticine“ nalezen v názvu knihy v použité literatuře, avšak v textu zmíněn není.

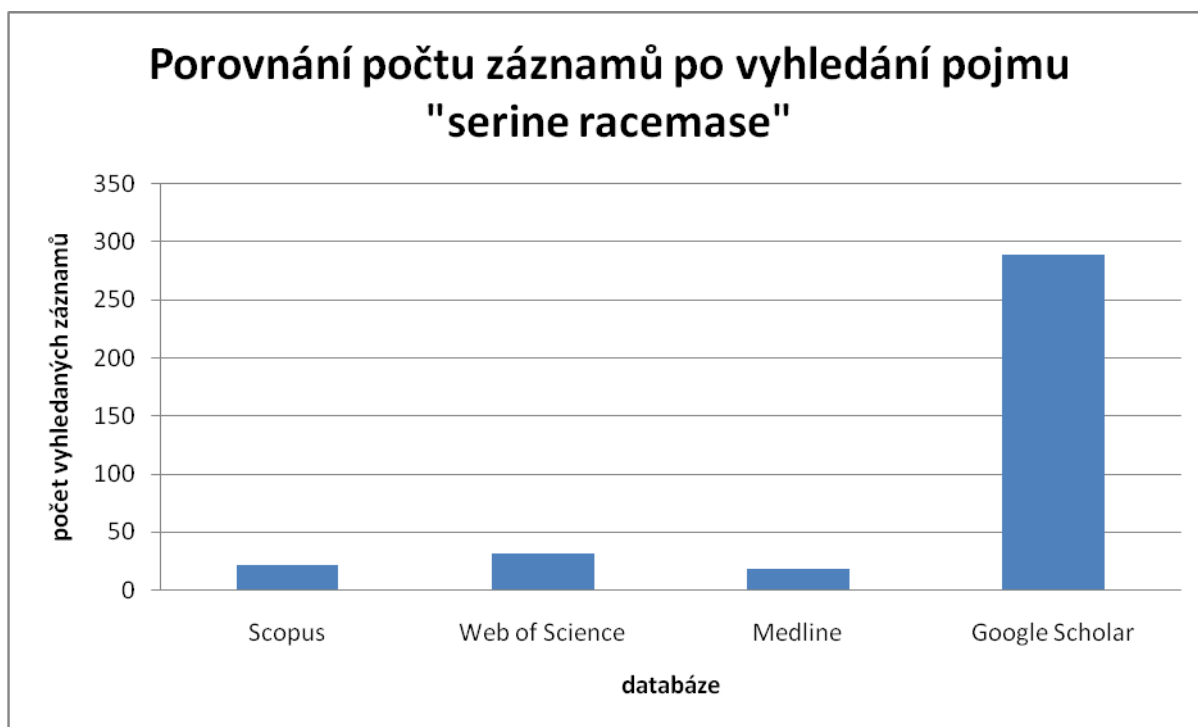
Jednou ze zajímavých vlastností databáze Google Scholar je prohledávání knih. Ze 160 výlučných záznamů jich bylo 40 citací z různých knih a 120 článků. Majoritní část z vyhledaných odkazů tedy bylo články, ale ani oněch 40 kapitol z knih není zanedbatelným číslem (jedná se o 25% záznamů).

Poslední informaci, kterou bych ráda uvedla, je, že řazení textů u Google Scholar zhruba odpovídá cíli tvůrců (viz. Kap. 1.2.4.. zkráceně: nejrelevantnější články se objevují vždy nejvýše). Adekvátní články byly opravdu k nalezení nejvýše, zatímco záznamy na konci se již vyhledávaného pojmu téměř vůbec netýkaly. Je ovšem nutno podotknout, že toto tvrzení lze potvrdit jen u většiny článků, rozhodně ne u všech. Poslední ze sedmnácti relevantních článků jsem našla v polovině druhé stránky záznamů (cca stopadesáté místo). To tedy znamená, že se na řazení u databáze Google Scholar spolehnout nedá.

Ráda bych také uvedla celkový počet relevantních článků u všech vyhledaných záznamů v databázích: SCOPUS poskytl z 36 vyhledaných celkem 17 relevantních článků, Web of Science 16 z celkového počtu 30 vyhledaných článků, a u databáze MEDLINE, která našla celkem 18 záznamů, jsem vyhodnotila 15 jako relevantních.

3.2. Pojem „serine racemase“

Situace z pohledu počtu vyhledaných článků v jednotlivých databázích byla totožná jako při vyhledávání předchozím: nejvíc záznamů opět našla databáze Google Scholar, což dokazuje obrázek 14:



Obr.14: Porovnání počtu vyhledaných záznamů pro pojem „serine racemase“

Stejně jako v předchozím hledání, i u tohoto bylo hlavním cílem vyhodnotit především relevanci či nerelevanci vyhledaných záznamů. Zatímco u ellipticinu byla relevance shledána při charakteristice protinádorového agens, u serinové racemasy budou za relevantní považovány všechny články, které se jí alespoň okrajově dotýkají (tzn.pokud se v abstraktu nebo plném textu toto sousloví alespoň vyskytne).

Z 22 vyhledaných výsledků v databázi SCOPUS bylo dle mého zhodnocení 16 relevantních článků, z čehož již u osmi článků bylo jasné z nadpisu, že se studie bude týkat právě serinové racemasy. Další tři články byly nerelevantní a tři studie pojem „serine racemase“ v abstraktu nenabídly (některé dokonce nenabídly ani onen abstrakt), avšak vyskytoval se v EMTREE drug terms, díky nimž byly tyto články zařazeny do výsledků. Plné texty pro vyhodnocení relevance bohužel nebyly dostupné.

V databázi Web of Science bylo nalezeno 33 studií. Vzhledem k faktu, že jedna studie byla přítomna dvakrát, vyhodnocovala jsem 32 záznamů. Z těch jsem shledala 11

relevantních hned podle názvu a devět relevantních po prostudování abstraktu, případně plného textu. Oproti databázi SCOPUS zde bylo ale velké množství záznamů neadekvátních, a to sedm. Dalších pět záznamů nešlo vyhodnotit, pojem nebyl přítomen v abstraktu a k plnému článku nebylo možné se dostat.

Zajímavý výsledek poskytla databáze MEDLINE. Z celkového počtu 18 záznamů se totiž ukázalo celkem 16 relevantních (u 11 to bylo zřejmé z nadpisu, u dalších pěti po pročtení abstraktu či hledání v plném textu) a pouze dva články nešly z tohoto hlediska vyhodnotit (nepřítomnost pojmu v abstraktu, plný text nedostupný). Z těchto čísel je jasné patrné, že žádný z článků vyhledaných v databázi MEDLINE nebyl nerelevantní.

Jak již bylo řečeno, databáze Google Scholar po zadání dotazu poskytla 289 výsledků. I v tomto případě jsem se snažila vyhodnotit relevanci daných článků dle výše zmíněných kritérií – relevantní jsou články, kde se pojem alespoň vyskytne. Po zhodnocení vyhledaných záznamů jsem shledala relevantních 79 článků a kapitol knih, dalších osm článků jsem považovala za relevantní, ačkoliv zadaný pojem byl přítomen v jiné grafické podobě (jednalo se o slova „ser-ine racemase“). Ačkoliv jsem ze zadání vyloučila vyhledávání záznamů s citacemi (tedy že zadaný problém se nevyskytuje např. v názvu knihy v použité literatuře), bylo dalších 5 záznamů takto vyhledáno a dané záznamy tedy byly označeny jako nerelevantní. Ostatní zbylé články byly vyhodnoceny jako nerelevantní (viz Tab.1).

Badatel často prohledávající databáze by si v tomto případě zadávání mohl pomoci uvozovkami – tedy zadat pojem v uvozovkách a databáze vyhledá články, kde by měla být zadaná fráze. Takovéto hledání je ovšem nebezpečné tím, že jsou vyhledány jen články s přesnou frází a jakákoliv její obměna je už problémem a program článek nevyhledá. Tento nedostatek ukazuje následující příklad: U jednoho z článků je v textu psáno „serine and alanine racemase“ – článek se tedy serinové racemasy bude týkat, avšak při zadání pojmu v uvozovkách není nalezen. U databázi SCOPUS, Web of Science a MEDLINE však, když připustíme výše popsání riziko, toto vyhledávání spolehlivě funguje, opravdu jsou zobrazeny články, kde daná fráze přesně je.

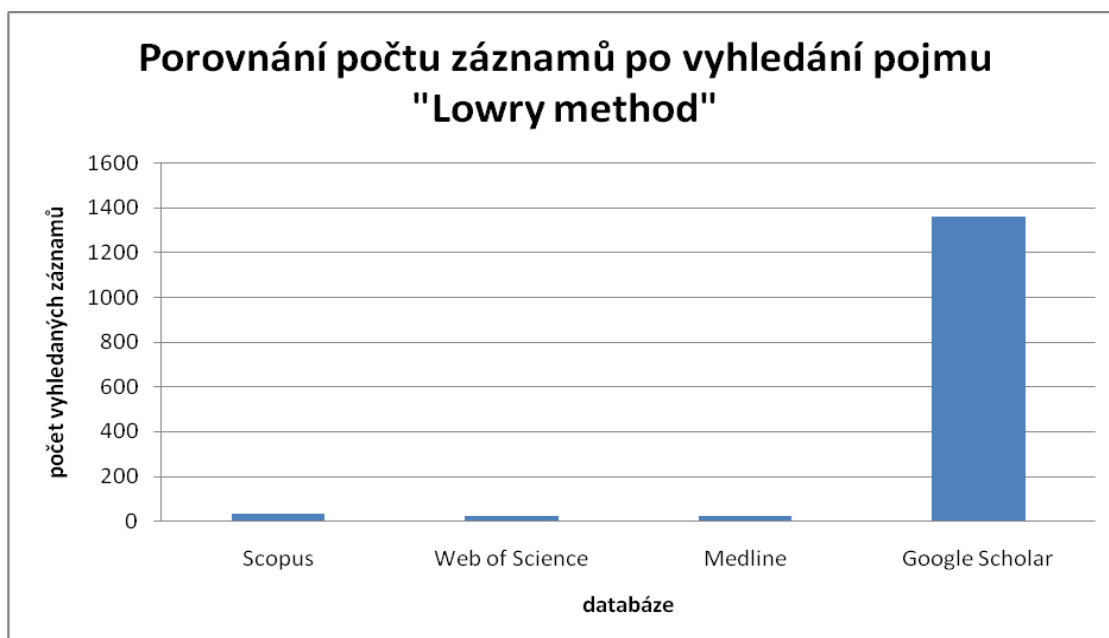
U databáze Google Scholar toto zadání však, zdá se, nefunguje přesně. Zadáme-li do vyhledávacího pole „serine racemase“ včetně uvozovek vyhledá se v databázi 101 záznamů. Několik prvních je relevantních hned na první pohled, pojem se vyskytuje již v nadpisu, druhá polovina vyhledaných záznamů už tak spolehlivě relevantní není. Po

otevření některých vyhledaných textů se stalo, že hledaný pojem opravdu přítomen v článku nebyl. Toto zadání je tedy třeba brát u databáze Google Scholar s nadhledem.

Ačkoliv tedy na první pohled vyšla databáze Google Scholar z tohoto hodnocení velmi dobře, ukázala se opět jako poměrně nespolehlivá. Relevantních článků je tedy zde velké množství, ale jak už bylo řečeno u hledání předchozích – vědec se při hledání relevantních článků musí probrat vysokým množstvím záznamů nezajímavých, protože nemalý počet relevantních článků se objevuje například až na druhé straně s vyhledanými záznamy ze tří.

3.3. Pojem „Lowry method“

Výrazně nejvíce článků pro tento výraz bylo opět vyhledáno v databázi Google Scholar.



Obr.15: Vyhodnocení vyhledávání pojmu „Lowry method“

Rozdíl mezi počtem vyhledaných pojmů je obrovský, ale je také nutno vyhodnotit, zda našel Google Scholar články adekvátní, případně jestli nastavení vyhledávače bylo správné a mezi vyhledané články se doopravdy nezahrnuly články autora Lowryho.

U vyhledaných článků jsem hodnotila relevanci. Jako relevantní jsem označila články, jež například modifikují postup u této biochemické metody, či studují její citlivost. Za nerelevantní naopak považuji záznamy, které pouze informují o využití Lowryho metody pro stanovení proteinů a dále ji nijak nehodnotí ani nemodifikují.

U databáze SCOPUS bylo vyhledáno celkem 36 článků, což značí druhý nejvyšší vyhledaný počet záznamů. U dohledatelných plných textů jsem tyto prostudovala, u nedostupných jsem podle hodnocení (viz výše) rozhodovala o relevanci z abstraktů. 26 článků bylo podle mého systému hodnocení nerelevantních – studie byla vyhledána pouze proto, že je v ní využito Lowryho metody pro stanovení proteinů, někdy naznačen postup práce, ale více článek neobsahuje. Pouze jeden článek byl plně relevantní – je zde popsáno zdokonalení Lowryho metody. Dalších 6 studií se týkalo porovnání Lowryho metody s jinými možnostmi určení proteinů (Bradfordové test, HPLC, biuretová metoda a další). Dále se vyskytly ještě tři záznamy, o kterých nelze říci, zda se jedná o studie relevantní či nerelevantní – ve všech případech nebyl k dispozici plný text, pouze abstrakt. U těchto dvou případů se v abstraktu hovoří o využití „modifikované Lowryho metody“, avšak není naznačeno, v čem modifikace spočívá. Pro správné vyhodnocení by tedy musel být k dispozici plný text a až poté by bylo možno zhodnotit, zda je toto zdokonalení popsáno a článek je tedy relevantní.

Navzdory mému předpokladu, že Web of Science by měla poskytnout alespoň stejné množství záznamů jako SCOPUS, tato databáze jich poskytla pouze 25. Pouze jeden záznam byl relevantní (mimo jiné to byl týž jako u předchozího hledání v databázi SCOPUS), pět jich porovnávalo různé metody určení proteinů (z čehož byly nalezeny čtyři stejné články jako ve SCOPUSu, jeden byl jiný) a 19 nalezených záznamů bylo označeno jako nerelevantních. U jedné studie jsem opět nemohla určit, zda se jedná o relevantní článek (opět jeden ze tří neurčitelných článků z hledání ve SCOPUSu).

Stejný počet vyhledaných záznamů (25) poskytla taktéž databáze MEDLINE. Největší množství článků bylo opět nerelevantních (19), dále dva záznamy relevantní (jeden z nich byl nalezen i ve SCOPUSu), tři články se zabývaly porovnáním biochemických metod (opět stejné články jako ve SCOPUSu). I v MEDLINE se objevil článek (stejný jako u SCOPUSu i Web of Science), u kterého z důvodu nepřístupnosti plného textu nebylo možné určit relevanci. Z výsledků poskytnutých databází MEDLINE je tedy patrné, že relevantní články byly stejné jako ve SCOPUSu a Web of Science, avšak celkově jich bylo o něco méně.

Google Scholar vyhledal přibližně 1360 článků a kapitol knih. U této databáze nebylo možné projít každý z vyhledaných záznamů zvlášť, zaměřila jsem se tedy na to, zda byly vyhledány články, které byly cílem (o Lowryho metodě, ne autora Lowryho). Po tomto hledání mohu říci, že záznamy se dotýkaly opravdu pouze Lowryho metody, žádný článek Lowryho jako autora nalezen opravdu nebyl. Na první pohled se záznamy

relevantní zdají, avšak po hlubší analýze by tomu tak být nemuselo. Pro ověření správnosti tohoto faktu by však byla třeba další porovnání. Při hledání se ukázalo, že databázi Google Scholar sice bylo udáno, že se našlo 1360 článků a kapitol knih, avšak zobrazit bylo možno pouze prvních 1000 (u zbylých 360 nelze nalézt ani název, ani abstrakt, není dostupná žádná informace). Zkoušela jsem otestovat databázi zadáním jiného dotazu, zda se tato skutečnost vyskytne i u jiných hledání a skutečně tomu tak bylo. Databáze tak pravděpodobně nutí člověka upřesnit vyhledávací dotaz, aby se s počtem výsledků vešel do čísla 1000 záznamů.

4. Diskuse

Nejdůležitější a časově nejnáročnější částí mé práce bylo vyhodnocení relevance článku. Výsledky zkoumání jsou shrnuty v následující tabulce 1:

Tab.1: shrnutí relevance vyhledaných výsledků

vyhledávaný pojem	databáze	celkový počet článků	celkový počet relevantních článků	celkový počet nerelevantních článků	články, u kterých relevance nebyla jasná
ellipticine	SCOPUS	36	17	7	12
	Web of Science	30	16	7	7
	MEDLINE	18	15	2	1
	Google Scholar	203	30	68	105
serine racemase	SCOPUS	22	16	3	3
	Web of Science	32	20	7	5
	MEDLINE	18	16	0	2
	Google Scholar	289	87	202	0
Lowry method	SCOPUS	36	1	26	3
	Web of Science	25	1	19	1
	MEDLINE	25	2	19	1
	Google Scholar	1360 ^a	Nebylo detailně hodnoceno		

^a dostupných pouze 1000

U prvního vyhledávání (pojem „ellipticine“) tedy sice našla největší počet relevantních článků databáze Google Scholar, avšak uživatel by tyto články musel hledat v celkovém počtu 203 záznamů. Druhý největší počet relevantních záznamů poskytla databáze SCOPUS, kdy 17 relevantních záznamů sice hledáme ve více než třiceti studiích (36), avšak čas hledáním strávený nebude takovým omezením jako hledání adekvátních výsledků u Google Scholar. Ani Web of Science, ani rozhraní PubMed databáze MEDLINE nevyšly z tohoto hledání nikterak špatně, databáze MEDLINE dokonce poskytla 83% adekvátních článků ze všech zde vyhledaných záznamů.

U zadání vyhledávacího dotazu „serine racemase“ našla největší počet relevantních článků databáze Web of Science. V tomto hledání vykázaly dobré výsledky všechny tři hodnocené databáze – SCOPUS malým počtem nedohledatelných plných

textů, Web of Science největším počtem adekvátních článků a MEDLINE díky faktu, že nebyl nalezen ani jeden nerelevantní záznam.

Při třetím hledání (pojmu „Lowry method“, které bylo ztíženo rozšířeným vyhledávacím dotazem) jsem vyhodnotila jako nejlepší databázi MEDLINE. Z 25 vyhledaných záznamů poskytla dva adekvátní, zatímco například u SCOPUSu byl nalezen menší počet relevantních článků a ještě byly hledány ve větším počtu vyhledaných záznamů. Pro databázi Google Scholar nebyla provedena detailnější analýza, neboť byly za relevantní považovány jiné články než v databázích předchozích.

Kromě počtu vyhledaných záznamů a zhodnocení jejich relevance lze také bibliografické databáze porovnat z hlediska několika dalších faktorů:

Tím prvním je uživatelský komfort („User-friendliness“). Po seznámení se s každou databází není jejich užívání obtížné, SCOPUS a Web of Science mají rozhraní dokonce velmi podobná, podle mého názoru nejjednodušší. PubMed má sice více možností omezení zadávaného výrazu (vymezení pohlaví či věku zkoumaného objektu a další), nicméně je zde obtížnější zadávání roku vydání. Například při vyhledávání pojmu „Lowry method“ zde bylo zadání složitější než u SCOPUSu či Web of Science, nicméně vedlo ke správnému cíli. Pro uživatele podstatným negativem je u PubMedu nepřítomnost historie hledání, či složky, kam by se ukládaly vybrané záznamy. Dále u PubMedu chybí rozřazení článků podle oborů, avšak jelikož jsem sama tento problém nemusela řešit, nevím jak velkou je toto překážkou. Google Scholar je z prvního pohledu uživatelsky velice jednoduchý, avšak chybí zde několik vyhledávacích možností. S tou první, podstatnou, jsem se sama setkala při vyhledávání pojmu „Lowry method“ – chybí zde možnost vyloučit jednu část hledaného pojmu pouze v některé části článku – v mém případě to byla nemožnost vyloučit slovo „Lowry“ pouze z autorů. Výsledky si při nejednoznačném zadání nemůže být uživatel jist, a tak před samotným hledáním informací o dané problematice musí nejprve zjistit, zda databáze opravdu vyhledala požadované články. Dalším negativem je, stejně jako u MEDLINE, absence složky s uloženými záznamy, či nepřítomnost historie hledání.

Dalším faktorem při výběru databáze může být to, zda databáze je placená, či neplacená a rozdíly mezi nimi. Jak již bylo řečeno, prvním důležitým kladem u databází placených je jednoduchost vyhledávání. Další, pravděpodobně nejdůležitější pozitivum, je lepší přístup k plným textům. U stejného článku, vyhledaného databázemi SCOPUS a Google Scholar, byl přístupný plný text pouze ze SCOPUSu. Google Scholar, jako

databáze neplacená, v tomto porovnání nedopadla dobře. MEDLINE se jeví jako lepší volba při nedostupnosti placené databáze.

Nyní se pokusím shrnout výsledky mé práce ve stěžejních bodech. Na začátek uvádím hlavní výhody a nevýhody každé z databází:

Tab.2: hlavní pozitiva a negativa databází

Databáze	hlavní pozitiva	hlavní negativa
SCOPUS	jednoduchost hledání, vysoký počet vyhledaných i adekvátních článků, možnost dodatečného omezení i po vyhledání záznamů	některým uživatelům se může zdát uspořádání výsledků nepřehledné
Web of Science	jednoduchost hledání, vysoký počet adekvátních článků	oproti SCOPUSu (druhá placená databáze) obtížnější rozšířené hledání
MEDLINE	vysoký počet adekvátních článků na počet článků celkově vyhledaných	nevyznačené hledané pojmy ve vyhledaných záznamech
Google Scholar	velké množství článků, což například pro zjištění základních informací o problému stačí	velké množství článků, ve kterých je obtížné vyhledat ty podstatné

Nejefektivnější databází pro biochemika, který potřebuje získat informace o určité problematice, je databáze SCOPUS. Je totiž uživatelsky velmi přijatelná, rychlá, přesnost vyhledaných záznamů je vysoká a oproti databázi MEDLINE, která se jevila také velice přijatelně, má ještě jednu vlastnost, která usnadní práci z pohledu času: ve vyhledaných záznamech i abstraktech samotných jsou zvýrazněna hledaná slova. Jde tedy snáze a rychleji posoudit, zda je pro mě článek podstatný či nikoliv. Databázi MEDLINE jsem kladně vyhodnotila z důvodu, že sice nevyhledá nijak vysoké množství záznamů, avšak naprostá většina z nich je relevantní. Databáze Web of Science je sice s databází SCOPUS co se relevantních článků týče srovnatelná, při tvorbě složitějšího spojení v rozšířeném hledání je však Web of Science obtížnější. Další negativum, které jsem na databázi shledala, byla nižší rychlost vyhledávání.

5. Závěr

Výsledkem porovnávání databází SCOPUS, Web of Science, MEDLINE a Google Scholar bylo zjištění, že pro tři dané biochemické problémy je nejméně výhodná databáze Google Scholar, další tři databáze jsou téměř srovnatelné. Z těchto tří poskytla nejlepší výsledky placená databáze SCOPUS, při nedostupnosti předplacených databází by velice dobře posloužila i MEDLINE s rozhraním PubMed. Pro optimální vyhledávání je dobré zkombinovat obě databáze. Při přístupu na placené „weby“ je vhodné využít i Web of Science, která poskytuje také relevantní záznamy (podíl relevantních ku vyhledaným článkům je velmi podobný databázi SCOPUS), avšak ve srovnání se SCOPUSEm a MEDLINE není tak efektivní (orientace v databázi, její rychlost).

Databáze SCOPUS má velmi přehledné rozhraní, ve kterém není obtížné se orientovat. Poskytla současně vysoký podíl relevantních článků.

Databáze MEDLINE (v mé práci s využitím rozhraní PubMed) je specializovaná spíše na medicínu než na biochemii či alespoň chemii. I přesto však podíl relevantních článků byl vysoký, avšak počet záznamů vyhledaných byl nižší než u ostatních databází.

Součást Web of Knowledge, databáze Web of Science, neposkytla špatné výsledky, pokud jde o podíl relevantních článků, byly výsledky srovnatelné s databází SCOPUS. V databázi SCOPUS však byla jednodušší orientace a stejně tak rychlost vyhledávání zde byla vyšší než u Web of Science.

Databáze SCOPUS tedy Web of Science předčila rychlostí a snazší orientací (relevance článků je srovnatelná), MEDLINE má zase jako neplacená databáze vyšší procento relevantních výsledků.

Databázi Google Scholar bych doporučila spíše při hledání základních informací o specializovaných problémech. Při nedostupnosti databází předplacených bych doporučila využít spíše MEDLINE. Program Google Scholar sice nalezne velké množství relevantních článků, ale jejich hledání mezi záznamy nepodstatnými je velice časově náročné. Při hledání zde je tedy třeba obrnit se trpělivostí.

6. Slovníček základních pojmů (dle TDKIV¹⁾)

Apriorní vztah = vztah mezi pojmy, resp. výrazy, který existuje nezávisle na větném kontextu¹⁾

Booleovská logika = logika založená na tom, že logické výroky jsou buď pravdivé nebo nepravdivé. Proměnné mohou mít jen jednu ze dvou hodnot (pravda, nepravda), vztahy mezi nimi se vyjadřují Booleovskými operátory.¹⁾

Booleovský operátor = operátor pro manipulaci s logickými hodnotami. Vyjadřuje logický vztah mezi dvěma vyhledávacími výrazy nebo prvky vyhledávání.

K nejpoužívanějším patří AND (logický součin), OR (logický součet) a NOT (negace).¹⁾

Databáze = soubor datových objektů v elektronické formě uložených společně podle jednoho tématu a zpřístupňovaných počítačem.¹⁾

Deskriptor = lexikální jednotka thesauru užívaná při indexaci k označení určitého pojmu v dokumentu nebo rešeršním požadavku.¹⁾

Thesaurus = řízený a měnitelný slovník deskriptorového jazyka uspořádaný tak, že zachycuje apriorní vztahy mezi lexikálními jednotkami.¹⁾

Uživatelské rozhraní = místo, kde dochází k výměně informací mezi člověkem a zařízením či programem¹⁾

Vyhledávací engine = jeden ze základních typu vyhledávání na internetu. Vyhledávací engine tvoří čtyři základní části: roboty (jejichž úkolem je sběr informací na internetu), indexační program (zpracovávající informace získané roboty), vyhledávací program (vyhledávací algoritmus a související programy) a grafické rozhraní (sbírá dotazy od uživatele, předává je vyhledávacímu stroji a zobrazuje výsledky hledání uživateli).¹⁾

7. Seznam použitých zkratek

Zkratky níže uvedené se týkají databází prohledávaných v článku Courbeta a Guyoda, odkud jsou jejich názvy přejaté⁴⁾

ATSDR – Agency Toxic Substances and Disease Registry

CCRIS – Chemical Carcinogenesis Research Inf Sys

CPDB – Carcinogenic Potency Database

ECOTOX – ECOTOXicology database

ECOTOXNET – Extension Toxicology Network

GAP – Genetic Activity Profiles

GENETOX – Genetic Toxicology

HCD – Hazardous Chemical Database

HSDB – Hazardous Substances Data Bank

HSFS – Hazardous Substance Fact Sheets

IARC – International Agency for Research on Cancer

ICSC – International Chemical Safety Cards

IPCS-INCHEM – International Program on Chemical Safety

IRIS – Integrated Risk Information Systems

NIOSH-IDLH – NIOSH Immediately Dangerous Life Health

NIOSH-PG – NIOSH Pocket Guide

NIOSH-DOE – NIOSH Department of Energy

NTP – National Toxicology Program Database

OSHAPEL – Occupational Safety and Health Administration Permissible Exposure Limits

SOLVEDB – Solvent-DB

TELETOX – Pesticides Toxicology Database

TOX-IN – Acute Human Intoxications Database

8. Seznam použité literatury:

- 1) *Databáze národní knihovny ČR* [online]. 2005 [cit. 2010-05-25]. TDKIV - Česká terminologická databáze knihovnictví a informační vědy . Dostupné z WWW: http://sigma.nkp.cz/F/?func=file&file_name=find-b&local_base=ktd
- 2) *Knihovny ČVUT* [online]. 2002 [cit. 2010-03-19]. Druhy databází. Dostupné z WWW: http://knihovny.cvut.cz/vychova/vychova2/databaze/druhy_databazi.html
- 3) Biarez, O.; Sarrut, B.; Doreau, C. G.; Etienne, J.: Comparison and evaluation of nine bibliographic databases concerning adverse drug reaction. *The Annals of Pharmacotherapy*. 25(1991), 1062-1065.
- 4) Guerbet, M.; Guyodo, G.: Efficiency of 22 Online databases in the search for Physiochemical, Toxicological and Ecotoxicological Information on Chemicals. *The Annals of Occupational Hygiene*. 46(2001), 261-268.
- 5) Guerbet, M.; Guyodo, G.: Comparison of 35 Electronic Databases for Environmental Risk Assessment. *The Annals of Occupational Hygiene*. 2001.
- 6) Hood, W. W.; Wilson, C. S.: Overlap in Bibliographic Databases. *Journal of the American Society for Information Science and technology*. 54(2003), 1091-1103.
- 7) McDonald, S.; Taylor, L.; Adams, C.: Searching the right database. A comparison of four databases for psychiatry journals. *Health Libraries Review*. 16(1999), 151-156.
- 8) Wilkins, T.; Gillies, R. A.; Davies, K.: EMBASE versus MEDLINE for family medicines searches : Can MEDLINE searches find the forest or a tree?. *Canadian Family Physician*. 51(2005), 848-849.
- 9) *EMBASE* [online]. 2010 [cit. 2010-05-12]. What is EMBASE?. Dostupné z WWW: <http://www.info.EMBASE.com/what-is-EMBASE/coverage>
- 10) Goodman, D.; Deis, L.: Web of Science (2004 version) and SCOPUS. *The Charleston Advisor* [online]. 6(2005) [cit. 2010-03-10]. Dostupný z WWW: http://charleston.publisher.ingentaconnect.com/search/article?title=Web+of+science&title_type=tka&year_from=1998&year_to=2009&database=1&pageSize=20&index=4
- 11) *SCOPUS* [online]. 2010 [cit. 2010-02-18]. SCOPUS in details. Dostupné z WWW: <http://info.SCOPUS.com/SCOPUS-in-detail/facts/>
- 12) *Elsevier* [online]. 20.2.2010 [cit. 2010-02-27]. EMTREE. Dostupné z WWW: <http://www.elsevier.com/wps/find/bibliographicdatabasedescription.authors/707574/description#description>

- 13) *Web of Knowledge* [online]. 2009 [cit. 2010-02-27]. Web of Science Database Selection. Dostupné z WWW:
http://apps.isiknowledge.com/additional_resources.do?highlighted_tab=additional_resources&product=WOS&SID=N2fBC3MADODFNNgLfC&cacheurl=no
- 14) *Web of Knowledge* [online]. 2009 [cit. 2010-02-27]. Web of Science Help. Dostupné z WWW:
http://images.isiknowledge.com/WOKRS49B3/help/WOS/h_database.html
- 15) Archambault, E.; Campbell, D.; Gingras, Y.; Lariviere, V: Comparing bibliometric statistics obtained from the Web of Science and SCOPUS. *Journal of the American Society for information Science and technology* [online]. 60(2009) [cit. 2010-03-14]. Dostupný z WWW:
<http://www3.interscience.wiley.com/cgi-bin/fulltext/122322374/HTMLSTART>
- 16) *U. S. National Library of Medicine: National Institute of Health* [online]. 2010 [cit. 2010-03-06]. PubMed Fact sheet. Dostupné z WWW:
<http://www.nlm.nih.gov/pubs/factsheets/PubMed.html>
- 17) AiP [online]. 2008 [cit. 2010-03-20]. MEDLINE. Dostupné z WWW:
<http://www.aip.cz/titul.php?titul=904>
- 18) *U. S. National Library of Medicine : National Institute of Health* [online]. 1.4.2010 [cit. 2010-05-02]. Medical Subject Headings (MeSH). Dostupný z WWW:
<http://www.nlm.nih.gov/pubs/factsheets/mesh.html>
- 19) *Tax payer access* [online]. 2008 [cit. 2010-03-08]. U.S. Congress Supports Public Access to Research. Dostupné z WWW:
<http://www.taxpayeraccess.org/supporters/us-congress/index.shtml>
- 20) *Tax payer access* [online]. 2009 [cit. 2010-03-08]. National Institutes of Health Public Access Policy. Dostupné z WWW:
<http://www.taxpayeraccess.org/issues/nih/index.shtml>
- 21) Google Scholar In *Wikipedia : the free encyclopedia* [online]. St. Petersburg (Florida): Wikipedia Foundation, , [cit. 2010-02-26]. Dostupné z WWW:
http://en.wikipedia.org/wiki/Google_Scholar
- 22) *Google Scholar* [online]. 2010 [cit. 2010-02-26]. Nápořád služby Google Scholar. Dostupné z WWW:
<http://scholar.google.cz/intl/cs/scholar/help.html>

- 23) An Examination of Citation Counts in a New Scholarly Communication Environment. *D-Lib Magazine* [online]. 11(2005) [cit. 2010-03-06]. Dostupný z WWW: <http://dlib.org/dlib/september05/bauer/09bauer.html>
- 24) Vine, R.: Google Scholar. *Journal of the Medical Library Association* [online]. 94(2009) [cit. 2010-02-27]. Dostupný z WWW: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1324783/?log%24=activity>
- 25) *Internet Tutorials* [online]. 10.5.2010 [cit. 2010-05-13]. Boolean Searching on the Internet. Dostupné z WWW: <http://www.internettutorials.net/boolean.asp>
- 26) *The Digital Object Identifier System* [online]. 6.5.2010 [cit. 2010-05-13]. The DOI System. Dostupné z WWW: www.doi.org

9. Příloha

Níže uvedené články v seznamu jsou záznamy nalezené v databázích SCOPUS, Web of Science a MEDLINE pro pojem „ellipticine“. Za názvem lze (pro případ zájmu o nějaký z článků) nalézt nejčastěji identifikátor DOI (digital object identifier, sloužící k identifikaci záznamů v digitálním prostředí²⁶⁾). Pokud nebyl článek označen DOI identifikátorem, využila jsem identifikátory jednotlivých databází: CODEN pro SCOPUS, IDS number pro web of Science a PMID jako identifikátor u PubMedu.

- 1) Martin, A. E. et al: 2-Acetyl-2,3,4,9-tetrahydro-1H-carbazol-1-one in the synthesis of heteroannulated carbazoles, DOI: 10.3184/030823407X272976
- 2) Sopkova-de Oliveira Santos, J et al.: 9-ethyl-1,4-dimethyl-6-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)-9H-carbazole and 6-bromo-9-ethyl-1,4-dimethyl-9H-carbazole, DOI: 10.1107/S0108270108020593
- 3) Kiran, Y. B. et al.: A green reagent for the iodination of phenols, DOI: 10.1055/s-2008-1078598
- 4) Xu, G. W. et al.: A high-content chemical screen identifies ellipticine as a modulator of p53 nuclear localization, DOI: 10.1007/s10495-007-0175-4
- 5) Mohanakrishnan, A. K. et al.: A one pot synthesis of annulated carbazole analogs, DOI: 10.1016/j.tetlet.2008.07.036
- 6) Su, J. G. et al.: Aldo-keto reductase 1C2 is essential for 1-nitropyrene's but not for benzo[a]pyrene's induction of p53 phosphorylation and apoptosis, DOI: 10.1016/j.tox.2007.11.022
- 7) Carroll, A. R. et al.: Alkaloids from the Australian rainforest tree *Ochrosia moovi*, DOI: 10.1021/np070655e
- 8) Gaddam, V. et al.: An efficient, one-pot synthesis of isomeric ellipticine derivatives through intramolecular imino-diels-alder reaction, DOI: 10.1021/ol800497u
- 9) Vogel, S. et al.: Aroyl hydrazones of 2-phenylindole-3-carbaldehydes as novel antimetabolic agents, DOI: 10.1016/j.bmc.2008.04.071
- 10) Berkaew, P. et al.: Aurocitrin and related polyketide metabolites from the wood-decay fungus *Hypocrea* sp. BCC 14122, DOI: 10.1021/np700740a
- 11) Kiralj, R. et al.: Comparative chemometric and QSAR/SAR study of structurally unrelated substrates of a MATE efflux pump VmrA from *V. parahaemolyticus*: Prediction of multidrug resistance, DOI: 10.1002/qsar.200630164

- 12) Thompson, D. et al.: Computer simulations reveal a novel nucleotide-type binding orientation for ellipticine-based anticancer c-kit kinase inhibitors, DOI: 10.1021/bi801239u
- 13) Johansson, I. et al.: Current research in drug metabolism, drug transport and drug development, DOI: 10.1358/dnp.2008.21.9.1308894
- 14) Purnaparte, K. et al.: Cytochrome P450s in the development of target-based anticancer drugs, DOI: 10.1016/j.canlet.2007.10.024
- 15) Potterat, O. et al.: Drug discovery and development with plant-derived compounds, DOI: 10.1007/978-3-7643-8117-2_2
- 16) Caruso, A. et al.: Efficient and simple synthesis of 6-aryl-1,4-dimethyl-9H-carbazoles, DOI: 10.3390/molecules13061312
- 17) Tian, E. et al.: Ellipticine derivative NSC 338258 represents a potential new antineoplastic agent for the treatment of multiple myeloma, DOI: 10.1158/1535-7163.MCT-07-0524
- 18) Fang, K. et al.: Ellipticine-induced apoptosis depends on Akt translocation and signaling in lung epithelial cancer cells, PMID: 18619705
- 19) Ishikura, M. et al.: Formal Synthesis of Olivacine via Indolylborate, IDS Number: 374NK
- 20) Abu-Sheaib, E. S. et al.: Heterocycles [h]-fused onto 4-oxoquinoline-3-carboxylic acid, Part VI . Synthesis and X-ray structure of model indolo[3,2-b]- and [2,3-b]pyrido[2,3-f]quinoxaline-3-carboxylic esters, DOI: 10.1007/s00706-008-0883-7
- 21) Moserova, M. et al.: Isolation and partial characterization of the adduct formed by 13-hydroxyellipticine with deoxyguanosine in DNA, IDS Number: 381RI
- 22) Wang, Y. et al.: Kaposi's sarcoma-associated herpesvirus ori-Lyt-dependent DNA replication: Involvement of host cellular factors, DOI: 10.1128/JVI.01319-07
- 23) Arlt, V. M. et al.: Metabolic activation of benzo[a]pyrene in vitro by hepatic cytochrome P450 contrasts with detoxification in vivo: experiments with hepatic cytochrome P450 reductase null mice, DOI: 10.1093/carcin/bgn002
- 24) Sabisz, M. et al.: Modulation of cellular response to anticancer treatment by Caffeine: Inhibition of cell cycle checkpoints, DNA repair and more, DOI: 10.2174/138920108785161497
- 25) Holt, P. A. et al.: Molecular docking of intercalators and groove-binders to nucleic acids using autodock and surflex, DOI: 10.1021/ci800063v

- 26) Szabó, L. F.: Molecular evolutionary lines in the formation of indole alkaloids derived from secologanin, IDS Number: 336LW
- 27) Cetinkol, O. P. et al.: Molecular recognition of poly(A) by small ligands: an alternative method of analysis reveals nanomolar, cooperative and shape-selective binding, PMCID: PMC2632892
- 28) Attia, S. M.: Mutagenicity of some topoisomerase II-interactive agents, CODEN: SPJOE
- 29) Lakshmi, V. M. et al.: N-demethylation is a major route of 2-amino-3-methylimidazo[4,5-f]quinoline metabolism in mouse, DOI: 10.1124/dmd.107.019166
- 30) Beghyn, T. et al.: Natural compounds: Leads or ideas? Bioinspired molecules for drug discovery, DOI: 10.1111/j.1747-0285.2008.00673.x
- 31) Miki, Y. et al.: Nitration of dimethyl 1-substituted indole-2,3-dicarboxylates: Synthesis of nitro- and aminoindole derivatives, IDS Number: 268FP
- 32) Lu, C. et al.: Non-genotoxic anti-neoplastic effects of ellipticine derivative NSC176327 in p53-deficient human colon carcinoma cells involve stimulation of p73, IDS Number: 392RA
- 33) Bae, S. I. et al.: PCNA damage caused by antineoplastic drugs, DOI: 10.1016/j.bcp.2008.09.003
- 34) Mirzayans, R. et al.: Pharmacological modulation of p53 function in cancer therapy, DOI: 10.2174/157436208785699721
- 35) Bunyapaiboonsri, T. et al.: Phenolic glycosides from the filamentous fungus *Acremonium* sp. BCC 14080, DOI: 10.1021/np070689m
- 36) Hernlund, E. et al.: Potentiation of chemotherapeutic drugs by energy metabolism inhibitors 2-deoxyglucose and etomoxir, DOI: 10.1002/ijc.23525
- 37) Lin, I. C. et al.: Predicting noncovalent interactions between aromatic biomolecules with London-dispersion-corrected DFT, DOI: 10.1021/jp0750102
- 38) Wang, W. et al.: Restoration of p53 to limit tumor growth, DOI: 10.1097/CCO.0b013e3282f31d6f
- 39) Stiborova, M. et al.: Role of hepatic cytochromes P450 in bioactivation of the anticancer drug ellipticine: Studies with the hepatic NADPH:Cytochrome P450 reductase null mouse, DOI: 10.1016/j.taap.2007.09.017
- 40) Pattamadilok, D. et al.: Seco-terpenoids and other constituents from *Elateriospermum tapos*, DOI: 10.1021/np070629g

- 41) Fung, S. Y. et al.: Sequence effect of self-assembling peptides on the complexation and in vitro delivery of the hydrophobic anticancer drug ellipticine, DOI: 10.1371/journal.pone.0001956
- 42) Stiborova, M. et al.: Studies with the hepatic NADPH: Cytochrome P450 reductase null mouse reveal the role of hepatic cytochromes P450 in bioactivation of the anticancer drug ellipticine, IDS Number: BGZ94
- 43) Tylinka, B. et al.: Synthesis and anticancer activity of new 1-substituted-6H-pyrido[4,3-b]carbazole derivatives, DOI: 10.1002/ardp.200700203
- 44) Abe, N. et al.: Synthesis and reactions of 3-ethynyl-2-(triphenylphosphino)-1-azaazulenes, IDS Number: 268FQ
- 45) Martinez-Esperon, A. F. et al.: Synthesis of carbazoles by dehydro Diels-Alder reactions ofynamides, DOI: 10.1016/j.tet-2008.02.029
- 46) Deveau, A. M. et al.: Synthesis of diketopiperazine-based carboline homodimers and in vitro growth inhibition of human carcinomas, DOI: 10.1016/j.bmcl.2008.05.022
- 47) Shoshan, M. C. et al.: Target specificity and off-target effects as determinants of cancer drug efficacy, DOI: 10.1517/17425255.4.3.273
- 48) Vazques, A. et al.: The genetics of the p53 pathway, apoptosis and cancer therapy, DOI: 10.1038/nrd2656
- 49) Jonsson, M. E. et al.: The zebrafish gill model: induction of CYP1A, EROD and PAH adduct formation., PMID: 19056132
- 50) Schmitt, M. et al.: Thermal C-2-C-6 cyclization of enyne-carbodiimides: Experimental evidence contradicts a diradical and suggests a carbene intermediate, DOI: 10.1021/jo701966h
- 51) Ramesh, N. et al.: Unusual dimerization of N-protected bromomethylindoles/benzyl bromide with arylmetal halides: generation of indolylmethyl/benzyl radical, DOI: 10.1016/j.tet.2007.12.047
- 52) Dey, A. et al.: Updates on p53: Modulation of p53 degradation as a therapeutic approach, DOI: 10.1038/sj.bjc.6604098
- 53) Shankar, J. et al.: UvrD helicase of Plasmodium falciparum, DOI: 10.1016/j.gene.2007.12.015
- 54) Li, X. J. et al.: Western-medicine-validated anti-tumor agents and traditional Chinese medicine, DOI: 10.1016/j.molmed.2007.11.002

Svoluji k zapůjčení této práce pro studijní účely a prosím, aby byla řádně vedena evidence vypůjčovateli.

[illegible]